VOL. 11 No. 8, pp. 221-252 FEBRUARY 1950

Philips Technical Review

DEALING WITH TECHNICAL PROBLEMS
RELATING TO THE PRODUCTS, PROCESSES AND INVESTIGATIONS OF
THE PHILIPS INDUSTRIES

EDITED BY THE RESEARCH LABORATORY OF N.V. PHILIPS' GLOEILAMPENFABRIEKEN, EINDHOVEN, NETHERLANDS

THE HELIX AS RESONATOR FOR GENERATING ULTRA HIGH FREQUENCIES

by B. B. van IPEREN.

621.385.1.029.6:621.396.615.143

It is a known fact that constructional difficulties are encountered when trying to make a triode suitable for ultra high frequencies. In recent years various types of valves have therefore been developed on entirely new principles. Remarkable results have been obtained with the travelling wave tube, a tube in which electromagnetic waves travel along a helix. This tube, which is particularly used for amplifiying ultra-short waves, can of course also serve for generating such waves. An oscillator based upon this principle is described in this article and its working is explained.

Introduction

In order to avoid, or at least greatly reduce, the difficulties arising when using triodes for ultra high frequencies [1])[2]), in recent years various valves have been developed which in this case could be used to replace the triode.

An example of such a type of valve is the induction tube [3]). Here, instead of striking an anode forming part of the resonant circuit, the electrons impart their energy inductively to the output circuit and are then taken up by an anode which may be of any size. This eliminates one of the drawbacks of the triode, namely that with the small dimensions essential for high frequencies only little power can be dissipated. But the difficulties with the input circuit remain the same.

A considerable improvement in this respect is the velocity-modulation valve [4]). In this valve a beam of electrons of a constant velocity passes a high-frequency alternating field. The electrons which thereby undergo a certain change in velocity, depending upon the moment at which they pass the field, then arrive in a field-free space where this velocity modulation is converted into a density modulation, so that the current receives an alternating-current component. Also with this valve the energy is given off inductively.

The two grids required for the velocity modulation may be of a coarser mesh in the case of a velocity-modulation valve than the control grid of the triode or the induction tube. This reduces the constructional difficulties. Since, however, the electrons remain a finite time in the modulating field (and also in the inductor fields) this gives rise to effects which have an unfavourable influence upon the working of the valve.

In the first place this finite transit time results in the action of the high-frequency electric field upon the electrons being less effective: the velocity modulation or the induced alternating current is less than when the transit time is negligible. In the second place there is a transit-time damping, the modulating field giving off energy to the beam.

It is possible to avoid the effect of the transit-time damping by giving the electrodes such a shape and placing them at such a distance that the damping becomes nil or even negative [5]. But then (if one keeps to the simple configuration of two grids or a gap) the velocity modulation or alternating current induction aimed at is diminished, so that in practice this remedy is not very attractive.

---

[1]) C. J. Bakker, Some characteristics of receiving valves in short-wave reception, Philips Techn. Rev. **1**, 171-177, 1936.

[2]) M. J. O. Strutt and A. van der Ziel, The behaviour of amplifier valves at very high frequencies, Philips Techn. Rev. **3**, 103-111, 1938. See also K. Rodenhuis, Two triodes for reception of decimetric waves, Philips Techn. Rev. **11**, 79-89, 1949 (No. 3).

[3]) This valve is described, i.a., by C. G. A. von Lindern and G. de Vries, Flat cavities as electrical resonators, Philips Techn. Rev. **8**, 149-160, 1946.

[4]) For a detailed description of the velocity-modulation valve see F. M. Penning, Velocity-modulation valves, Philips Techn. Rev. **8**, 214-224, 1946.

From this it is not to be deduced that a negative transit time damping is never applied. There is a valve in which this is utilized. The electron beam is then made to pass only one modulating field. If the distance of the electrodes has been well chosen then the beam can give off energy to the field and if this energy is more than is needed for generating the high-frequency electric field required then the system can oscillate. The valve referred to, which bears the name of monotron [6], works, however, only with a very high current, owing to the small absolute value of the negative transit-time damping. The small value of this damping is due to the fact that with the usual high-frequency fields the potential is a linear or at least monotone function of place. This will be reverted to later.

However, in the development of valves for very high frequencies a new principle is now being applied whereby use is made of an electric field the strength of which is a periodic function of place. The valves built according to this principle were at first considered for use as amplifiers [7][8] and were then given the name of "travelling wave tube". The simple construction of these valves makes them attractive also for generating ultra-short waves. In this article we shall particularly discuss the use of this valve as oscillator, but first of all a brief account will be given of its working as an amplifier.

Principle of the travelling wave tube used as amplifier

In a travelling wave tube it is ensured that the electro-magnetic wave that is to be amplified has a velocity, in a certain direction, which is only a fraction of the velocity in the free space. The wave thus retarded then has a reciprocal action with an electron beam.

There are various ways of retarding an electromagnetic wave, but theoretical investigations indicate that the application of a helix for this purpose offers the most favourable solution.

When an electromagnetic wave is introduced at one end of the helix it travels along the wire at a velocity approximating that of light. Viewed in the axial direction of the helix however the velocity is greatly reduced, in fact in the ratio of the pitch of the helix to its circumference. In this way it is possible to obtain along the axis of the helix a travelling electromagnetic wave with a velocity of, say, 1/10 of the velocity of light and with the electrical vector lying along the direction of the axis. When an electron beam is sent along the axis a reciprocal action is set up between the beam and the electromagnetic wave. This situation is diagrammatically represented in *fig. 1*. Calculation shows that in a certain range of electron velocities the beam may impart energy to the wave and thus amplify it.

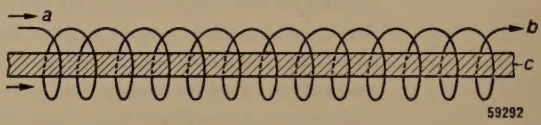


Fig. 1. Diagrammatic representation of the application of a helix with an electron current passing through it, used as amplifier. *a* = the arriving wave, *b* = the amplified wave, *c* = the electron current.

In broad terms this can be explained in the following way. The wave gives a velocity modulation to the beam and this is converted into a density modulation, so that the electron current in the beam receives an alternating-current component. The latter in turn yields energy to the wave and the amplified wave modulates the beam again, and so on. It appears that owing to this mechanism the amplitude of the wave increases exponentially with the distance travelled.

A closer investigation shows that the action of the valve is somewhat more complicated than this. The amplitude of a travelling wave as a function of the distance x and the time t can be represented by the formula

$$A(x,t) = A(0)\,e^{j(\omega t - \Gamma x)} \quad . \; . \; . \; . \; . \; . \; . \; . \quad (1)$$

where ω is the angular frequency and Γ the so-called constant of propagation. Now in the theory of the travelling wave tube it is assumed that all electrical quantities counting for this wave movement can also be represented in the manner indicated by formula (1). Calculating the reciprocal action between the electron beam and the field of the helix, we arrive at an equation of the fourth degree in Γ. Of the four roots of this equation two are real, one positive and one negative. This means that a wave travelling to the left and one travelling to the right can move along the tube with constant amplitude. The other two solutions for Γ are conjugate complex, which means to say that in the exponent of the e-power is a real term which is positive for one root and negative for the other.

---

[5] These effects will be dealt with later.

[6] The monotron has been fully discussed by J. J. Müller and R. Rostas in Helv. Phys. Acta 13, 435-450, 1940.

[7] R. Kompfner, Wireless World 52, 369-372, 1946.

[8] J. R. Pierce, Proc. Inst. Red. Engrs 35, 111-123, 1947.

These two solutions thus both correspond to a wave travelling to the right, one of which is amplified and the other attenuated. The amplification appears to reach the maximum when the velocity of the electrons is equal to the velocity which the wave would have along the axis of the helix if there were no beam.

An important property of the arrangement described here is the fact that the amplification is only very slightly dependent upon the frequency. Care must be taken, however, to ensure that only travelling waves do indeed occur; in other words, for all these frequencies there must not be any reflection at the end of the helix. If this condition is fulfilled then the band width — by which is to be understood the difference between the two frequencies at which the power amplification has dropped to one half — may amount to about one-third of the mean frequency.

The generation of oscillations

In the foregoing a brief description has been given of the manner in which the travelling-wave tube functions as amplifier. The usual way of turning an amplifier into an oscillator is to feed back part of the energy from the output side of the tube to the input side. With the tube in question however there is a much simpler method. By removing entirely the matched load that was essential when using the tube as an amplifier, then, owing to the reflection, standing waves are obtained at the ends of the helix, which then acts as a resonator.

This is quite comparable to other devices where both travelling and standing waves may occur, as for instance in a transmission line or a wave guide; there, too, a travelling wave is obtained if there is a matched load and a standing wave if there is no such matching. The electric field along the axis of the helix then likewise becomes a standing field, such that the amplitude varies along that axis roughly sinusoidally. As will be explained below, a beam of electrons with a suitably chosen velocity will then yield energy to that field. This can also be expressed by saying that in that case a negative transit-time damping arises, just as it does in a monotron. It appears, however, that with the sinusoidal field of the helix this negative transit-time damping may be much greater in absolute value than is the case with an approximately constant field, such as obtained with a monotron.

An oscillating "travelling wave tube" may also be regarded as a sort of velocity-modulating valve, but with a number of modulating fields and a number of inducing fields placed one behind the other, instead of one of each.

The question now arises as to whether the improvement due to the more favourable form of the field is not neutralized by the replacement of the high-impedance cavity resonators usually employed with velocity-modulation and similar valves by a helix, which is generally considered to be a much less effective resonator.

There is, however, the somewhat surprising fact that, as a calculation shows, the power required to generate a certain field in a part of a helix having a length of a half wavelength is practically equal to the power used in generating a field of the same size in a good cavity resonator. Thus the impedances of the circuits are approximately the same in both cases. This only holds, however, so long as there are so many waves on the helix that the radiation may be ignored.

Before going into these matters further, let us consider more closely the helix as a resonator.

Resonances of helices

The general problem of the field of a helix acting as the carrier of an electromagnetic wave is mathematically so difficult that no solution has yet been found for it.

For the case in which we are most interested, namely that of a helix which is long compared with its diameter, we may, however, to a good approximation say that the intensity of the current along the wire changes sinusoidally. Denoting the wavelength, measured along the wire of the helix, by λ_s and the total length of wire by L, we must have:

$$L = n\lambda_s, \qquad (n = \tfrac{1}{2}, 1, 1\tfrac{1}{2}, \dots).$$

Considering, however, the wavelength along the axis of the helix, λ_z, it is clear that it must equally hold that

$$l = n\lambda_z, \qquad (n = \tfrac{1}{2}, 1, 1\tfrac{1}{2}, \dots),$$

where l is the length of the helix (see *fig. 2*). Denoting the distance between two turns (the pitch of the helix) by h and the radius of a turn by R, we get the equation

$$\frac{l}{L} = \frac{\lambda_z}{\lambda_s} = \frac{h}{2\pi R} = \sin\varphi,$$

where φ is the angle made by the wire with a plane perpendicular to the axis of the helix. The electric field along the axis of the helix is likewise sinusoidal with the wavelength λ_z.

We now have to put the question what resonance frequencies correspond to the various values of n, thus what wavelength λ_0 in the free space corresponds to each value of n (or λ_z and λ_s respectively). So far it has only been found possible to calculate this for the idealized case of an infinitely long cylinder along the surface of which currents are only possible at a given angle to the axis of the cylinder. One can imagine such a cylinder as being built up of a large number of helically wound, very thin wires lying parallel to each other on the periphery of the cylinder. For this case an exact solution can be given. Lenz has already done this for very small values of φ [9]. The solution is also possible for arbitrary values of φ [8] [10], it thereby having been found that for all values of φ occurring in practice the result is practically independent of φ, We can therefore confine our considerations to the case where φ is small.

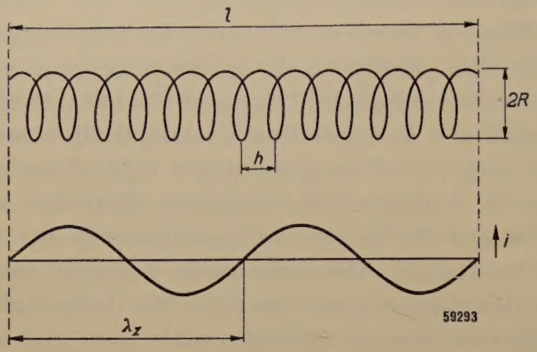


Fig. 2. In the case of a helix of which the length l is large compared with the diameter $2R$ the current flows along the wire sinusoidally. In the case drawn here $n = 2$. The wavelength along the axis is represented by λ_z.

The calculation then yields as result a relation between λ_s/λ_0 and $2\pi R/\lambda_z$. This relation is represented in *fig. 3* by the curve *I*. If λ_s/λ_0 were equal to 1 this would mean that a wave travels along the wire of the helix with the velocity of light. From the curve *I* it therefore appears that the velocity along the wire is indeed over a large area approximately the velocity of light [11] and that a considerable change takes place when λ_z is large compared with R.

Although the above-mentioned theory cannot, of course, take full account of the actual behaviour of a helix, in the case that is of most interest to us, namely that of a long helix oscillating in one of its higher resonance frequencies, there appears to be little divergence from the theory. This may be seen from curve *II* in fig. 3, representing the result of measurements taken on a helix with $l = 155$ mm, $R = 6$ mm, wire thickness $d = 0.7$ mm and $h = 4$ mm. The numbers given at the measuring points indicate what harmonic the point represents, thus the number of half waves on the helix.

It is to be noted that the measurements give much greater deviations from this theory when the fundamental wave of the helix is measured (thus for $n = \frac{1}{2}$). In this connection Drude [12] has carried out extensive measurements with helices of different dimensions. He found that the fundamental frequency depends but little upon d and h, but that it is particularly determined by the total length of the wire L and the ratio l/R. His results for a helix where $h/d = 2.4$ have been plotted in curve *III* for values of $l/(2R)$ varying from 6 to 0.05. The right-hand part of curve *III* thus relates to the range of short coils, for which the above-mentioned theory is no longer applicable.

For this range a theory has been worked out by Lenz [9] and Szasz [13], whilst later Hallén [14] gave an improved theory for the fundamental wave of short coils (curve *IV*) which agrees rather well with Drude's measurements (curve *III*). Curve *V*, theoretically derived by Lenz and Szasz, relates to the various resonances of a very short coil of a certain shape (where $l/(2R) = 0.122$). Finally curve *IV* gives two measuring points by these authors for such a coil.

Contrary to what has been found in the case of long helices, the velocity of the wave in the case of short coils is always less than the velocity of light.

Transmission of energy from an electron beam to an electric alternating field

From what has been found in the foregoing it follows that in the case of a helix we have to do with an axially directed electric alternating field which can be represented by the formula

$$E = E_0 \cos\left(\frac{2\pi n}{l}\, z\right) \cos \omega t \quad . \; . \; . \; . \; . \quad (2)$$
$$(-l/2 < z < +l/2)$$

Travelling along the axis is a beam of electrons the initial velocity of which we shall denote by v_0. The question is now what power this beam will yield to the field.

In general such a problem can only be solved by very laborious graphical and numerical methods, but in the exceptional case where the power transmitted is small compared with the power with which the electrons reach the alter-

---

[9] W. Lenz, Berechnung der Eigenschwingungen einlagiger Spulen, Ann. Phys. Leipzig, 43, 749-797, 1914.

[10] J. Chew and J. D. Jackson, Field theory of traveling wave tubes, Proc. Inst. Rad. Engrs 36, 853-863, 1948.

[11] This fact is closely related to the already mentioned property of the wide band width of travelling-wave amplifiers.

[12] P. Drude, Ann. Phys. Leipzig 9, 590-610, 1902.

[13] O. Szasz, Mathematischer Beitrag zu der Abhandlung des Hrn. Lenz, (see [9]), Ann. Phys. Leipzig 43, 789-809, 1914.

[14] E. Hallén, Ueber die elektrischen Schwingungen in drahtförmigen Leitern, Upsala Universitets-årsskrift 1930.

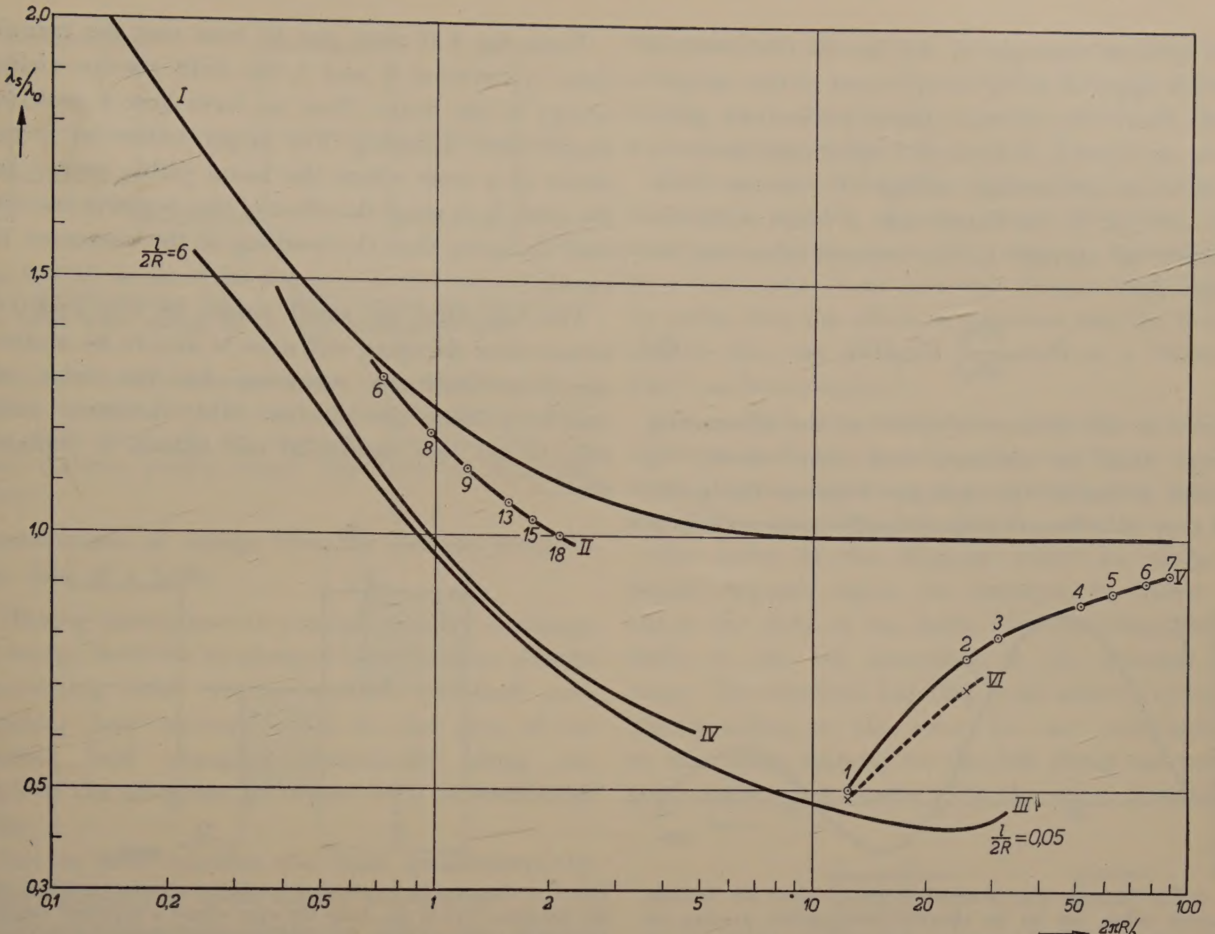


Fig. 3. The relation between the quantities λ_s/λ_0 and $2\pi R/\lambda_z$ (λ_0, λ_s and λ_z are the wavelengths respectively in the free space, along the wire and along the axis of the cylinder; R is the radius of the cylinder). *I* Theoretical curve according to Lenz for a helically conducting infinitely long cylinder. *II* Measurements with a helix of the length $l = 155$ mm and radius $R = 6$ mm, where the thickness of the wire was $d = 0.7$ mm and the spacing between the wires $h = 4$ mm. (The numbers at the measuring points denote the number of half waves on the helix.) *III* Measurements taken by Drude where only the fundamental wave was considered and with variations of $l/(2R)$ from 6 to 0.05. *IV* Theoretical curve according to Hallén for the fundamental wave ($n = \frac{1}{2}$). *V* Theoretical curve according to Lenz and Szasz for a short coil. *VI* Measurements taken by these authors with a coil with $l/(2R) = 0.122$.

nating field the solution can be given in a simple closed form, and for any arbitrary shape of the field along the path of the electrons [15]). One cannot, it is true, calculate in this way the yield that can be reached with a given system, but it is indeed possible to determine whether oscillations will arise with a given intensity of the electron current, and it can also be calculated at what electron velocity this will be the case.

The solution referred to is as follows. The electric field is represented by $E = E(z) \cos \omega t$, where $E(z)$ denotes the variation, as yet still arbitrary, of the electric field strength along the z-axis. It then appears that a passing electron receives from this field an amount of energy which to a first approximation is equal to:

$$\Delta_1 = e \cdot \left\} A(\xi) \cos\delta - B(\xi) \sin\delta \right\{ \quad . \; . \; . \; . \quad (3)$$

where $A(\xi) = \int_{-l/2}^{l/2} E(z) \cos \xi z \, dz$ and $B(\xi) = \int_{-l/2}^{l/2} E(z) \sin \xi z \, dz$.

Here $-l/_2$ and $l/_2$ are the limits of z beyond which $E(z) = 0$, δ is a parameter indicating at what instant the electron enters the field and $\xi = \omega/v_0$ (ω = angular frequency, v_0 = electron velocity), whilst e is the charge of the electron.

From formula (3) it can be determined what variation in velocity the electron undergoes in the field, thus what in the case of a velocity-modulation valve is called the velocity modulation. The total energy that the beam receives per second from the field is found by determining the mean value of Δ_1 over a cycle of the oscillation, thus over the area in which δ changes from 0 to 2π. This is found to be zero.

Considering the power transmission to a second approximation we find that the average is no longer zero. The formula then applying is:

$$\Delta_2 = -\frac{e^2\omega^3}{2m v_0} \left\} A(\xi) A'(\xi) + B(\xi) B'(\xi) \right\{ \quad . \; . \quad (4)$$

where $A'(\xi)$ and $B'(\xi)$ are derivatives with respect to ξ.

---

[15]) See B. B. van Iperen, On the generation of electromagnetic oscillations in a spiral by an axial electron current, Philips Res. Rep. **4**, 20-30, 1949 (No. 1).

To give an example of the use of this method we shall apply it to the simple case of the monotron. There the electron beam passes two grids which are spaced a distance l apart and between which is an alternating voltage $V_{\sim} \cos \omega t$. Now $E(z) = V_{\sim}/l$. It is found that a beam with the intensity of current i_0 per second takes up an energy P:

$$P = \frac{2 e i_0}{m v_0^2} V_{\sim}^2 \, \mathrm{f}(\tau), \quad . \; . \; . \; . \; . \quad (5)$$

where τ is the number of cycles of the alternating voltage that an electron with the velocity v_0 requires to travel the distance between the grids, and $\mathrm{f}(\tau)$ is a function graphically represented in *fig. 4*.

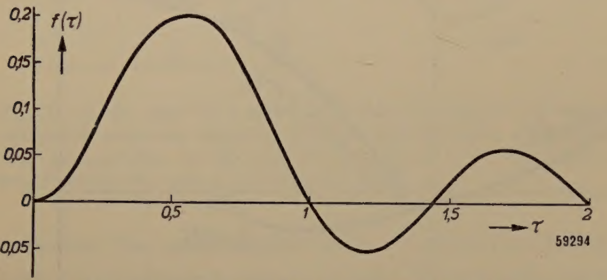


Fig. 4. The function $\mathrm{f}(\tau)$ to which is proportional the amount of power taken up by an electron beam when passing an electric alternating field between two grids. τ is the transit time, i.e. the number of cycles of the alternating voltage that an electron with a certain velocity needs to travel the distance between the grids.

If the impedance of the circuit coupling the grids is Z then the power dissipated therein is $V_{\sim}^2/Z$.

It is clear that oscillations can only occur if P is negative, and

$$|P| > V_{\sim}^2/Z \quad . \; . \; . \; . \; . \; . \; . \quad (6)$$

Introducing the acceleration voltage of the beam V_0 (for which $eV_0 = \frac{1}{2} m v_0^2$ applies) formula (6) becomes

$$-\frac{i_0}{V_0} \, \mathrm{f}(\tau) > \frac{1}{Z} . \quad . \; . \; . \; . \; . \; . \quad (7)$$

The quantity $Z_0 = V_0/i_0$ may be termed the beam impedance of the valve. With the aid of this we can express the condition for the occurrence of oscillations as

$$-\frac{1}{Z_0} \, \mathrm{f}(\tau) > \frac{1}{Z} . \quad . \; . \; . \; . \; . \; . \; . \quad (8)$$

Since for all values of τ the term $|\mathrm{f}(\tau)|$ is much smaller than 1, to comply with this condition Z_0 must be small. Consequently the current must have a high intensity.

From fig. 4 it may also be seen that for transit times τ between 0 and 1 the field always yields energy to the beam. Thus we have here a positive transit-time damping. For larger values of τ we arrive in a zone where the beam yields energy to the field. It is upon the effect of this negative transit time damping that the working of the monotron is based.

The fact that for small values of τ a positive transit-time damping will arise is also to be understood qualitatively, assuming for the sake of simplicity that the electric field variation with time is not just sinusoidal but follows a "square sine".

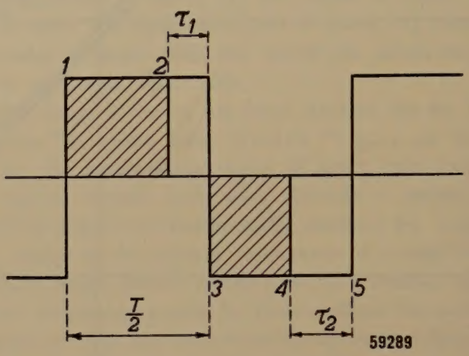


Fig. 5. A diagram for explaining qualitatively the variation of the function $\mathrm{f}(\tau)$ in fig. 4 for the case where τ is rather small. Here it is assumed that the field strength changes with time according to a "square sine". Since the transit time τ_1 of the accelerated electrons is smaller than that of the retarded electrons, τ_2, the number of electrons which are exclusively accelerated will be greater than the number of those which are only retarded. Thus on the average the beam takes up energy from the field.

All electrons entering the field at the instants *1* to *2* (see *fig. 5*) receive the same increase of energy, whilst all electrons arriving at the instants *3* to *4* undergo an equal reduction of energy. Since, however, the transit time τ_1 of the accelerated electrons is less than that of the retarded electrons, τ_2, the amounts of energy belonging to these two groups of electrons do not neutralize each other. The electrons arriving between *2* and *3* and remaining for a part of their transit time inside the retarding field are fewer in number than those which arrive between *4* and *5* and remain for a part of their time in the accelerating field. As a result, therefore, the beam takes up energy.

It is rather more difficult to understand that if τ is slightly greater than 1 the effect is the reverse, but this can be made plausible with the aid of *fig. 6*. We choose τ of such a value that the electron *1* reaching the field at the beginning of the accelerating needs just one cycle to pass the field. We can then choose a group of electrons entering between *1* and *2* in such a way that they leave the field in the area *1'-2'*. On the average these electrons take up energy from the field. In the area *2-3* (corresponding to *2'-3'*), however, there are

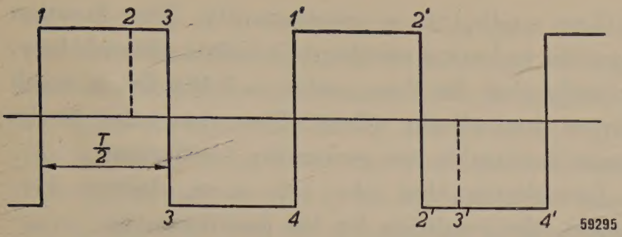


Fig. 6. With the aid of this diagram it is made plausible that for τ slightly greater than 1 on the average the electrons of the beam yield energy to the electromagnetic field.

electrons which yield energy to the field, whilst farther on all the electrons from the area *3-4* (corresponding to *3'-4'*) also yield energy. In the aggregate, therefore, there are more electrons yielding energy than there are dissipating energy.

Transmission of energy from an electron beam to the field of a helix

Having shown how in a simple case the exchange of energy between an electron beam and an electric alternating field can be studied, we shall now consider how matters stand in the case of an electric field changing sinusoidally along the path of the electrons, as occurs with an oscillating helix.

Let us first consider this case qualitatively before giving the result of the calculation.

The standing wave along the helix can always be imagined as being resolved into two travelling waves, one moving to the left and the other to the right. We assume that the electrons travelling to the right have a velocity v_0 which is approximately equal to the velocity of the two waves. Compared with the wave travelling to the right these electrons have a velocity which is small with respect to v_0, but the wave travelling to the left passes them with the velocity $2v_0$. Viewed from the electron, the latter wave causes the field strength to change very rapidly. Thus the effect of this upon the electron must be small, so that we can disregard the influence of this wave.

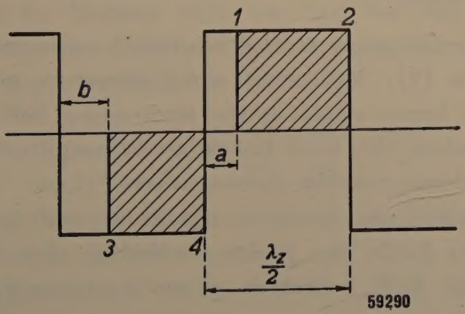


Fig. 7. The transfer of energy from the field of a coil to an electron beam. In this diagram the electric field along the axis of the helix is represented in a system of coordinates moving with the velocity of the wave. This field, which in the said system does not change with time, is actually sinusoidal but is represented here by a "square sine". With the aid of this diagram it is shown that on the average the electrons of the beam take up energy from the field when their velocity v_0 is slightly less than the velocity v_g of the wave on the helix. In that case the area *b* is greater than the area *a*.

To investigate the action of the other wave let us imagine a system of coordinates following the movement of this wave. The wave is represented by a sine curve which does not change with time. In order that the effect in question may be made clearer this sine is again replaced by a "square sine", as drawn in *fig. 7*.

Let us first assume that the velocity v_0 of the electrons is slightly less than the velocity v_g of the wave. In the beginning the electrons then travel along the wave to the left with the velocity $v_g - v_0$.

The point in the diagram where an electron begins depends upon the moment at which it enters the field of the helix. This starting point shifts to the left according to the moment of entry. The electrons entering in an interval of time corresponding to the area *1-2* are accelerated, so that their velocity to the left along the wave gets smaller and smaller. On the other hand the electrons beginning between *3* and *4* steadily increase in velocity to the left. The distance over which the electrons from the area *1-2* move to the left during their passage through the helix is therefore less than the distance covered by the electrons from the area *3-4* during their passage. The number of electrons dissipating energy during the whole of their passage is therefore greater than the number of electrons exclusively yielding energy to the wave, or, to put it in other words, the area *a* of the electrons passing over to the retarding field is smaller than the area *b* of the electrons passing over to the accelerating field. On the average, therefore, the electrons of the beam will take up energy from the field.

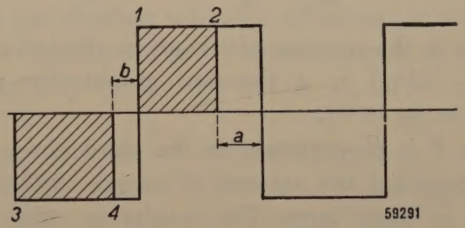


Fig. 8. Taken on the average the electrons of a beam will yield energy to the field when their velocity v_0 is a little higher than the velocity v_g of the wave. Now the area *b* is smaller than the area *a*.

Let us now suppose that the velocity v_0 of the electrons in the beam is slightly greater than v_g. In this case *fig. 8* applies. With respect to the wave the electrons now move to the right with the velocity $v_0 - v_g$. Those beginning between *1* and *2* are accelerated. We now find that the area *a* of the electrons passing over to the retarding field is larger than the area *b* of the electrons passing over to the accelerating field. Thus on the average the beam yields energy to the field, so that oscillations are generated.

From this reasoning it also follows that the desired effect can always be brought about to a sufficient degree, however small the electric field may be, by causing the electrons to move along with the wave for a sufficient length of time, thus by making the helix long enough.

The calculation of the exchange of energy for this particular case can again be carried out by the method already indicated. The alternating field arising along the axis of an oscillating helix is given by formula (2):

$$E(z) = E_0 \cos 2\pi n z/l\,.$$

In this case we find for the energy yielded per second by a beam with intensity of current i_0:

$$P = \frac{e\,i_0}{m v_0^3} \cdot \frac{E_0^2 l^3}{8}\,\mathrm{F}(2\pi n - \frac{\omega l}{v_0}) \quad . \quad . \quad (9)$$

where *n* is the number of waves on the helix and $F(2\pi n - \omega l/v_0)$ is a function graphically represented in *fig. 9*[16]).

Since *P* is proportional to the third power of *l*, by increasing *l* the amount of energy yielded can always be made large. The maximum value of *P* is obtained when $2\pi n - \omega l/v_0$ is so chosen that $\mathrm{F}(2\pi n - \omega l/v_0)$ is a maximum[17]). This function appears to have a number of maxima. One of these, namely that for $2\pi n - \omega l/v_0 = 0.42 \times 2\pi$, is much larger than all the others. Thus this value is the most favourable for generating oscillations.

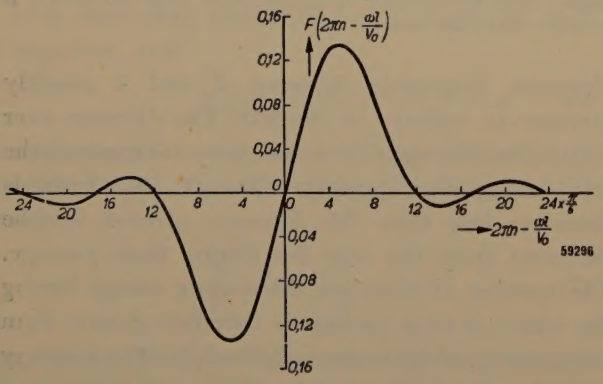


Fig. 9. Graph of the function $\mathrm{F}(\pi n - \omega l/v_0)$ to which is proportionate the power yielded by a passing electron beam to the sinusoidal electric field of a helix. The diagram shows that this function has several maxima, the greatest of which occurs at $2\pi n - \omega l/v_0 = 0.42 \times 2\pi$.

Considering that $\nu l = n\nu\lambda_z = n v_g$ (where $\nu = \omega/2\pi$), the condition for the maximum is:

$$n\frac{v_0 - v_g}{v_0} = 0.42, \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

from which it appears that v_g must indeed be smaller than v_0. For generating oscillations it is therefore necessary that the electrons travel slightly faster than the wave, and this difference in velocity has to be smaller the more waves there are on the helix.

Formula (10) can be reduced to a more practicable form by introducing the acceleration voltage V_0 and the wavelength λ_0, which then gives:

$$\frac{500\,l}{\lambda_0 \sqrt{V_0}} = n - 0.42 \quad (V_0 \text{ in volts}). \quad . \quad (11)$$

When we have a helix of given dimensions we can first determine from fig. 3 (curve *I*) the wavelength λ_0 corresponding to any value of *n* and then, with the aid of formula (11), calculate the optimum voltage V_0 required for generating these waves.

Formula (11) holds for the highest maximum of $\mathrm{F}(2\pi n - \omega l/v_0)$. A maximum about 10 times as small is found for $2\pi n - \omega l/v_0 = -1.16 \times 2\pi$. Oscillations can also be obtained for this value. Formula (11) then takes the form of:

$$\frac{500\,l}{\lambda_0 \sqrt{V_0}} = n + 1.16\,. \quad . \quad . \quad . \quad . \quad . \quad (11a)$$

Calculation of the current intensity at which oscillations begin

In the foregoing section we found an expression (formula (9)) indicating what power a passing electron beam yields to the alternating field of a helix when the field has a given amplitude E_0. We shall now consider the case where $\mathrm{F}(2\pi n - \omega l/v_0)$ has reached its greatest maximum and is thus equal to 0.13. The power yielded is then $P = i_0 e/8 m v_0^3 \cdot E_0^2 l^3$, which in a more practical form reads:

$$P = 26\,\frac{l^3}{\lambda_0 V_0^{3/2}} \cdot i_0 E_0^2 \text{ watts}\,. \quad . \quad . \quad . \quad (12)$$

On the other hand we have to know what power is required to maintain the alternating field E_0. This is difficult to calculate exactly but a good approximation can be obtained in the following way. Let us first assume that the wire forming the helix has the thickness 0 and that through this wire an alternating current flows the amplitude of which changes sinusoidally along the wire with the wavelength λ_s. From the current distribution thus given one can also (with the aid of the continuity equation) find the distribution of the charge along the wire, and from that, according to the well-known solution of Maxwell's equations, with a given distribution of current density and charge density in space, the electromagnetic field in the surroundings of the helix can be calculated. Now all that we are interested in here is the field along the axis. The exact formula for this is rather complicated [18]), but for nearly all cases occurring in practice the following approximative formula suffices:

---

[16]) This formula holds only to a good approximation if the number of waves on the helix is not too small say, at least 3.

[17]) This is not quite correct, because also the factor preceding *F* contains variable quantities. A correction for this, which is fairly small, is to be found in the article quoted in footnote [15]).

$$E_0 = 60\, I_0 \frac{\lambda_0}{R^{1/2}\,\lambda_z^{3/2}} \cdot e^{-\frac{2\pi R}{\lambda_z}} \quad (\mathrm{V/cm}) \quad . \quad (13)$$

where I_0 is the amplitude of the current in the wire.

The next step is to calculate the power dissipated in the wire as a consequence of the alternating current I_0 flowing through it. Of course we must in the first place take into account the fact that owing to skin effect the current flows only along the surface of the wire. Actually there is also the fact that the current is not uniformly distributed over the circumference of the wire but flows for the greater part along the side of the wire facing the axis of the helix, but here no account has been taken of this effect. This having been disregarded, the calculated losses will be smaller than the actual losses.

By taking these calculated losses as being equal to the energy yield of the beam per second as given by formula (12) we find for the current intensity at which oscillations begin (for the case of a copper helix):

$$i_0 = 3.8 \cdot 10^{-9}\, \frac{V_0^{3/2}}{n^3} \cdot \frac{RL}{d\,\lambda_0^{3/2}}\, e^{\frac{4\pi R n}{l}}, \quad . \quad . \quad (14)$$

where d is the thickness and L the total length of the wire. In this formula i_0 and V_0 are expressed in amperes and volts, and L, l, d, λ_0 and R in centimetres.

Some experiments for checking the calculations

The easiest way to produce an oscillator working on the principle described in the foregoing is to take a vacuum tube of silica (e.g. 20 cm long and 15 mm thick) having a cathode system at one end and an anode at the other end, and to wind a copper wire round it. In order to avoid any divergence of the electron beam this tube has to be placed in the magnetic field of a coil.

But such an arrangement is not suitable for checking the validity of the formulae derived above, because owing to charges on the wall of the tube electrostatic fields are set up which affect the velocity of the electron beam. To overcome this difficulty we used a tube with a helix (of molybdenum wire with a cross section of 0.7 mm) placed inside it instead of around it. With the helix oscillating at a resonance frequency we then determined at what tube voltage the oscillations with the respective frequency were strongest. The corresponding value of n, the number of waves on the helix, was found in a simple manner by sliding along the helix a small incandescent lamp, which glowed strongest where there was a maximum field strength.

According to formulae (11) and (11a), $1/(\lambda_0\sqrt{V_0})$ when plotted as a function of n should give a straight line. The experiment was first carried out with the smallest intensity of current at which the tube oscillated. *Fig. 10* shows (curve *2*) that the points measured do indeed lie on a straight line. Measurements taken with a slightly stronger current (15 mA) appear to give a straight line slightly lower (curve *3*). Line *1* in fig. 10 represents the theoretical curve corresponding to the lower range of oscillation according to formula (11a). Since in that case oscillations could only be produced with a much stronger current, only one point of this curve could be measured. For the value corresponding to this point the test appeared to be in agreement with the theory.

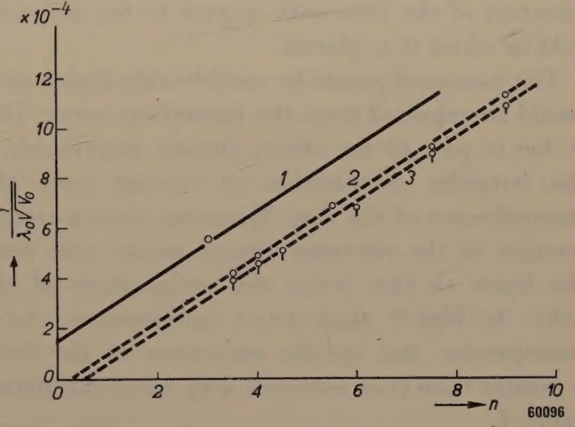


Fig. 10. The relation between $1/\lambda_0\sqrt{V_0})$ and n. (1) the theoretical curve for the lower range of oscillations according to formula (11a). (*2*) the measured points, which indeed appear to lie on a straight line. (*3*) measurements taken at a slightly stronger current (15 mA).

---

[18]) See E. Roubine, L'onde électr. **27**, 203-208, 1947.

A second test was carried out to check the validity of formula (14) indicating at what intensity of current the oscillations begin. The results have been plotted in *fig. 11*. Curve *1* has been calculated from formula (14). Curves *2* and *3* are the results of two series of measurements. The fact that these do not coincide better is due to the value of i_0 being very sensitive to any small deviation of the direction of the tube with respect to the magnetic field in which it is placed.

The measured points lie considerably higher than would be expected from the theoretical curve. This is due in part to the effect, already mentioned, of the irregular distribution of current over the circumference of the wire. Moreover, since a certain fraction of the electrons always passes over from the beam to the helix, the temperature of the helix is higher than room temperature. As a consequence the specific resistance of the helix is greater than that reckoned with when calculating curve *1*.

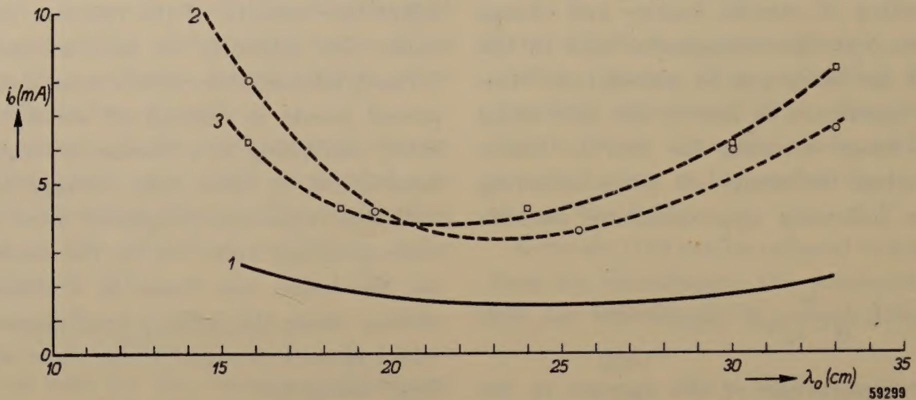


Fig. 11. The theory concerning the current intensity at which oscillations begin, checked by measurements. Curve *1* has been calculated from formula (14). Curves *2* and *3* have been plotted from two series of measurements.

Focusing of the electron beam by means of positive ions

It has already been pointed out that it is necessary to ensure that the electron beam does not diverge, and it has been said that this can be done with the aid of a magnetic field. The coils needed for this purpose, however, make the apparatus less easy to handle, though the tube itself is quite simple. Now it has been known for some time [19]) that the mutual repulsion of the electrons in a beam can be compensated or at least greatly reduced by introducing a certain quantity of some gas or other into the tube. The gas atoms are then ionized by the beam and owing to the presence of positive ions, which remain in the beam a fairly long time, the negative space charge of the electrons is neutralised. As a consequence the beam shows less tendency to diverge.

It has indeed been found possible, with the aid

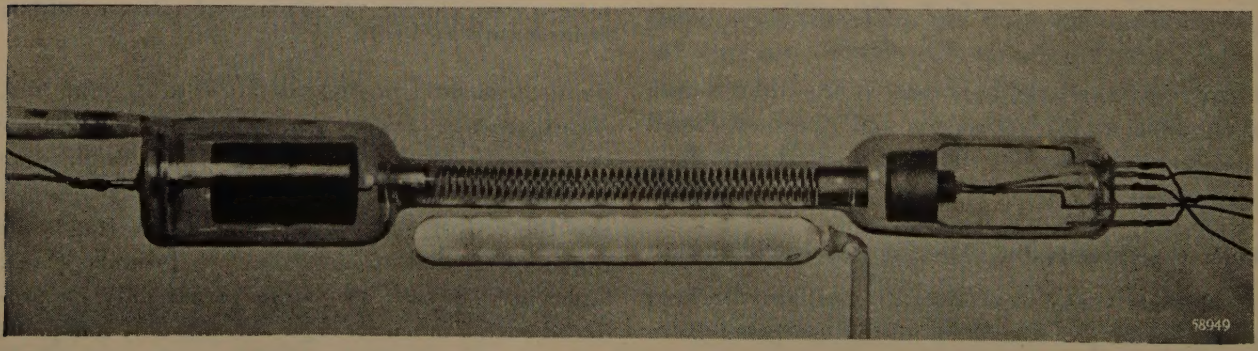


Fig. 12. Experimental model of a tube with gas-concentrated beam. The standing waves on the helix have been made visible with the aid of a neon tube. (Total length of the helix 12 cm; wavelength 20 cm.)

---

[19]) J. B. Johnson, Phys. Rev. 17, 420-421, 1921.

of a suitable electron gun and a rare gas of a certain pressure (for instance neon of 10^{-3} to $2 \cdot 10^{-3}$ mm Hg), to cause the tube to oscillate without employing focusing coils. The electron gun used gives a somewhat convergent beam, which beyond the focal point, if the tube were not filled with gas, would diverge again. However, owing to the presence of the gas ions the divergence is very much more gradual, so that the major part of the electron current is still able to reach the anode.

Fig. 12 is a photograph of an oscillating tube. It also shows how the standing waves on the helix can easily be made visible, by means of a small tube filled with neon (pressure about 10 mm Hg) placed alongside the helix; the neon glows strongly at the maxima of the electric field strength.

Summary. In the application of triodes for ultra high frequencies difficulties arise owing to the transit time of the electrons no longer being small compared with the cycle of the high-frequency alternating voltage. These difficulties are only partly overcome by using an induction tube, a velocity-modulation valve or a monotron. Matters are further improved by using a travelling-wave tube. This has already been studied in its application as amplifier. It also provides, however, a simple means of generating short waves. In this tube an electromagnetic wave travels along a coiled wire, whilst an electron beam is sent along the axis of the coil. After showing how this system can be used as an amplifier, the mechanism of the oscillation of the tube is discussed. The resonances of the coil play an important part in this mechanism. It is demonstrated that with a travelling wave tube it is a fairly simple matter to calculate under what conditions oscillations are possible and at what intensity of current they begin. Finally the results are given of some experiments carried out which confirm the accuracy of the theory given. For practical purposes it is of importance that the magnetic field for focusing the electron beam can be dispensed with when the tube is filled with a rare gas under a certain pressure.

HEATING BY MEANS OF HIGH-FREQUENCY FIELDS

II. CAPACITIVE HEATING

by M. STEL and E. C. WITSENBURG *).

621.364.16:679.562:674.24

The phenomenon of dielectric losses has been known for a long time, but its application as a source of heat (capacitive heating), on an industrial scale, dates back to only just before 1940. In the years following 1940 there has been a great advance in capacitive heating. This is connected with the rapid development in plastics, in the manufacture of which capacitive heating has in many cases led to a considerable saving and improvement of quality. Another important field of application is the wood-working industry, in particular the manufacture of plywood in flat and bent shapes.

Recently an article was published in this Journal dealing with induction heating 1). By this is understood the heating of conductive materials by means of electric currents induced therein by a magnetic alternating field. We shall now deal with capacitive heating, the heating of non-conductive materials by means of the dielectric losses occurring therein when they are placed in an electric alternating field. Whereas these losses often constitute an undesired property in insulators, here they form the basis of the heating process.

Nature of the dielectric losses

Very briefly we shall recall here some of the main features of dielectric losses.

As already discussed at length in this Journal 2), dielectric losses are the result of after-effect phenomena. When the electrical field strength in a dielectric for instance is increased suddenly, but after that kept constant, then the dielectric displacement likewise makes a jump but instead of then remaining constant it continues to increase gradually until it approaches a final value. In simple cases this "after-effect" takes place according to an exponential function of time. The relaxation time (after-effect time) τ is in many cases in essence a diffusion time, shorter or longer according to whether charge-carriers or dipoles in the dielectric have more or less freedom of movement or are able to change their direction under the influence of the changing field strength.

If the field strength alternates sinusoidally with time, as is the case with the capacitor in *fig. 1a* connected to an alternating voltage, then the after-effect results in a phase shift δ between the field strength and the dielectric displacement. Accordingly the current flowing through the capacitor leads by an angle φ slightly less than 90° ($\varphi = 90° - \delta$). Corresponding to the phase shift δ are certain losses, viz. the dielectric losses, and that is why δ is also called the loss angle.

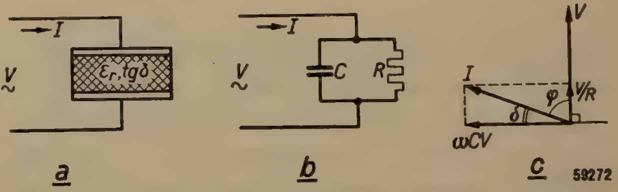


Fig. 1. *a*) Dielectric (with the relative dielectric constant ε_r and the loss angle δ) between two electrodes connected to a source of alternating voltage V. *b*) Equivalent circuit for the capacitor represented in (*a*). *c*) Vector diagram. The current I indicated in (*a*) and (*b*) has a component ωCV in quadrature with V and a component V/R in phase with V.

In an equivalent circuit (fig. 1*b*) these losses can be taken into account by imagining a resistance R shunted across a loss-free capacitor (with the capacitance C of the capacitor represented in fig. 1*a*). The corresponding vector diagram in fig. 1*c* shows that, with V representing the R.M.S. value of the alternating voltage applied and ω the angular frequency, the power W converted into heat amounts to

$$W = \frac{V^2}{R} = \omega C V^2 \cdot \frac{1}{\omega C R} = \omega C V^2 \tan\delta \,. \quad (1)$$

For the simple cases mentioned above it can be calculated (see the article quoted in footnote 2)) that tan δ as a function of ω has a maximum at $\omega = 1/\tau$. Measurements taken on most dielectrics show, however, that in a wide frequency band tan δ is only little dependent on ω. This is due to the fact that the moving charge-carriers or dipoles do not all behave in the same way, which means to say that one has to do with not one but a series of after-effect times τ. It is not difficult to understand that as a consequence the curve representing $\tan\delta$ as a function of ω is much flatter than when there is only one after-effect time to be taken into account.

---

*) Philips Telecommunication Industry, formerly N.S.F., Hilversum (Netherlands).

1) E. C. Witsenburg, Heating by high-frequency fields, I. Induction heating, Philips Tech. Rev. **11**, 165-175, 1949 (No. 6).

2) J. L. Snoek and F. K. du Pré, Several after-effect phenomena and related losses in alternating fields, Philips Techn. Rev. **8**, 57-64, 1946.

Something similar is the case with the dielectric constant, which, likewise due to the large number of after-effect times, in most materials changes but little with the frequency.

The heat generated by the dielectric losses will now be considered quantitatively.

Capacitive heating of a homogeneous medium

If the electric field in the capacitor of fig. 1*a* may be regarded as being homogeneous (thus ignoring boundary effects) then formula (1) may be written in the following form:

$$W = 0.556\cdot 10^{-10}\cdot(\varepsilon_r \tan\delta)\cdot E^2 f\cdot\frac{M}{\varrho}\ \text{watts}, \quad (2)$$

where ε_r is the (relative) dielectric constant of the dielectric to be heated, E the field strength therein (R.M.S. value in V/m), f the frequency (in c/s), M the mass (in kg) and ϱ the density (in kg/m³) of the dielectric.

Eq. (2) is derived from eq. (1) as follows: Denoting the surface area of the plates by A and the distance by s, the capacitance is [3]):

$$C = \frac{\varepsilon A}{s} = \frac{\varepsilon_0\varepsilon_r A s}{s^2},$$

where ε = the absolute and ε_r the relative dielectric constant, ε_0 = the dielectric constant of the vacuum (in the Giorgi system $= 8.855 \times 10^{-12}$ F/m).

Considering that As = the volume of the dielectric $= M/\varrho$ and that $V/s = E$, one easily arrives at formula (2).

In order to save time one will try to make the generated power W so great that the final temperature desired is reached very quickly. In that case, for a rough calculation, we may disregard the loss of heat during the short heating period, the more so since there is little conduction and radiation: little conduction because the material is a poor electrical and thus also a poor thermal conductor, and little radiation because with capacitive heating there are usually no higher temperatures than 200 °C.

Disregarding, therefore, the small loss of heat, we have:

$$W = Mc\frac{\mathrm{d}\vartheta}{\mathrm{d}t}\ \text{watts}, \quad .\ .\ .\ .\ (3)$$

where c = the specific heat (in joules per kg and per °C), ϑ = the temperature (in °C) and t = the time (in seconds).

From (2) and (3) it follows that

$$\frac{\mathrm{d}\vartheta}{\mathrm{d}t} = 0.556\times 10^{-6}\cdot\frac{\varepsilon_r \tan\delta}{\varrho c}E^2 f\quad .\ .\ .\ (4)$$

As we have seen, for most materials ε_r and $\tan\delta$ show little dependency upon the frequency: $(\varepsilon_r \tan\delta)/\varrho c$ can then be regarded as a material constant (apart from temperature dependency). For a given material formula (4) gives the relation between the field strength, the frequency and the rate of increase of temperature.

If it is asked, for instance, what frequency is required to heat a piece of wood ($\varepsilon_r = 3$, $\tan\delta = 6\times 10^{-2}$, $\varrho = 600$ kg/m³, c = 1700 J/kg × °C) to 90 °C in $1\frac{1}{2}$ minutes, then with $E = 10^5$ V/m we find from (4): $f = 10^7$ c/s = 10 Mc/s.

As regards the field strength, 10^5 V/m = 100 V/mm is about the limit to which one can go without risk of breakdown. If vapours arise from the object during the heating process these may condense on the electrodes and thus increase the risk of breakdown, in which case it is advisable to choose a weaker field strength. In order to maintain the same rate of heating then, according to (4), a higher frequency is needed.

Capacitive heating of a non-homogeneous medium

In fig. 1*a* we started from a homogeneous dielectric filling the entire space between the plates of the capacitor. In practice, when gluing wood for instance, one often has to do with a charge that is not homogeneous. We shall therefore now consider two simple cases: a charge having two different homogeneous dielectrics, where the interface layer is parallel to the direction of the field strength, and the case where it is perpendicular thereto.

The first case is represented diagrammatically in *fig. 2a*; it occurs, inter alia, in the gluing together

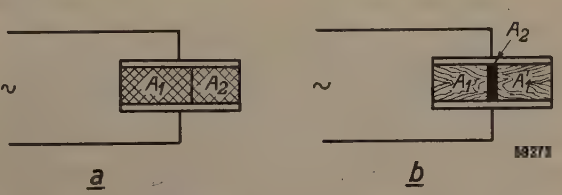


Fig. 2. *a*) Two dielectrics (A_1, A_2) connected in parallel. *b*) Example of parallel connection when gluing with capacitive heating. A_1, A_1' are blocks of wood with a layer of glue (A_2) between them. The glue is heated more quickly than the wood.

---

[3]) See for instance P. Cornelius, The rationalized Giorgi system with absolute volt and ampere as applied in electrical engineering, Philips Techn. Rev. **10**, 79-86, 1948 (No. 3).

of two pieces of wood (fig. 2*b*). Electrically speaking we have here two capacitors connected in parallel. For each of the dielectrics formula (4) applies, with the same values of E and f but with different values for ε_r, tan δ, ϱ and c. Distinguishing the dielectrics with the indices 1 and 2, we have:

$$\frac{\mathrm{d}\vartheta_1}{\mathrm{d}t} : \frac{\mathrm{d}\vartheta_2}{\mathrm{d}t} = \frac{\varepsilon_1 \tan\delta_1}{\varepsilon_2 \tan\delta_2} \cdot \frac{\varrho_2\, c_2}{\varrho_1\, c_1} . \quad . \; . \; . \quad (5)$$

Somewhat less simple is the case of a stratified dielectric where the layers are perpendicular to the direction of the field strength, such as the layers of wood and glue in the manufacture of plywood. In *fig. 3* this case is illustrated for two layers, electrically connected in series.

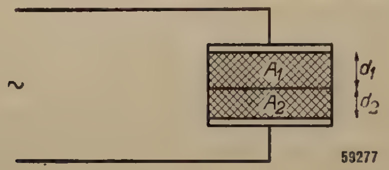


Fig. 3. Two dielectrics (A_1, A_2) in series.

As opposed to the case of fig. 2, the field strengths in the different materials are now unequal. In the layer with the dielectric constant ε_1, the loss angle δ_1 and the thickness d_1 the field strength E_1 is:

$$E_1 = \frac{\varepsilon_2}{\varepsilon_1 d_2 + \varepsilon_2 d_1} V, \; . \; . \; . \; . \; . \quad (6a)$$

and in the other layer the field strength E_2 is:

$$E_2 = \frac{\varepsilon_1}{\varepsilon_1 d_2 + \varepsilon_2 d_1} V \; . \; . \; . \; . \; . \quad (6b)$$

For the ratio of the rates of heating we find from (4):

$$\frac{\mathrm{d}\vartheta_1}{\mathrm{d}t} : \frac{\mathrm{d}\vartheta_2}{\mathrm{d}t} = \frac{\varepsilon_2 \tan\delta_1}{\varepsilon_1 \tan\delta_2} \cdot \frac{\varrho_2\, c_2}{\varrho_1\, c_1} , \quad . \; . \; . \quad (7)$$

which differs from (5) only in the fact that here ε_1 and ε_2 are interchanged. It is this change which accounts for the somewhat surprising fact that when two materials are "connected in parallel" the ratio of the rates of heating may be entirely different from that in the case of "series connection" of the same materials. When, for instance, two pieces of wood are glued together with urea-formaldehyde glue in the manner according to fig. 2*b*, then the glue is heated much quicker than the wood (selective heating). With the "series connection" (fig. 3) on the other hand the heating of the glue is only a little quicker than that of the wood.

For gluing, selective heating is highly favourable, because it is mainly a question of the heating of the glue; the heat generated in the wood is useless. By choosing a certain position and shape of the electrodes one can often promote selective heating, as illustrated in *fig. 4*, where a layer of veneer has to be glued, over a large area, onto a thicker layer of wood. This is a situation that does not lend itself to ordinary "parallel connection". In order to get a better selective heating than is possible with "series connection" the electrodes here have been made in the form of parallel and alternately poled strips applied only on the side of the layer of veneer. As is the case with "parallel connection", here the lines of force run more or less parallel to the interface layer.

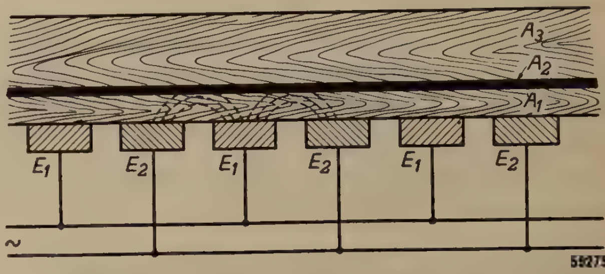


Fig. 4. A thin layer of veneer A_1 has to be glued onto a thick plank A_3 with a glue A_2. The electrodes E_1, E_2 are applied and connected in such a way that the lines of force (dotted arcs) in the glue are roughly parallel to the interface between the glue and the wood; the same effect (selective heating) is then obtained as with parallel connection.

Voltage distribution along the workpiece

In the foregoing it has been tacitly assumed that the potential is the same at all points of each of the electrodes. This, however, holds to a good approximation only so long as the wavelength corresponding to the working frequency is large with respect to the largest dimension of the electrodes. If the wavelength is comparable with that dimension (owing to the workpiece being large or the frequency being high) the field strength in the workpiece will vary from point to point, so that there will be an irregular distribution of heat.

With wood and similar materials tan δ is so small that to a first approximation the wave phenomenon at the electrodes can be regarded as a standing wave. If V_0 is the potential difference (R.M.S. value) between the points of the electrodes where the input wires are connected, then the potential difference V_l between two other points of the electrodes at the distance l from the input points is:

$$V_l = \frac{V_0}{\cos\left(\frac{l}{\lambda}\,2\pi\right)}. \qquad \ldots\ldots\ldots \quad (8)$$

So long as $l \ll \lambda$, $V_l \approx V_0$, but when l becomes comparable to λ or $\gg \lambda$ then in certain areas V_l will be much greater than V_0.

If, for instance, a heat distribution is permitted where the extreme temperatures differ no more than 5% from the mean, thus where the lowest temperature is not more than 10% lower than the highest temperature, then the smallest field strength must not be less than 95% of the largest field strength. From (8) we then find for the maximum distance $l_{\max}$ between the edge of the dielectric and the input point:

$$l_{\max} = \frac{\text{arc}\cos 0.95}{2\pi}\,\lambda = 0.05\,\lambda.$$

For λ we can write: $\lambda = v/f = c_l/(f\sqrt{\varepsilon_r})$, where v = velocity in the dielectric and c_l the velocity in vacuum $\approx 3 \times 10^8$ m/sec. Thus

$$l_{\max} \approx \frac{15 \cdot 10^6}{f\sqrt{\varepsilon_r}}\ \text{m}. \quad \ldots\ldots \quad (9)$$

With, for instance, $f = 15 \cdot 10^6$ c/s and $\varepsilon_r = 3$ (wood) we find: $l_{\max} \approx 0.6$ m. Thus, with a temperature tolerance of ± 5%, this frequency is suitable for heating sheets of wood of a maximum length of 1.2 m (current input in the middle of the electrodes).

Generators for capacitive heating

As may have been seen from the article quoted in footnote [1]), in the case of induction heating the choice of the frequency is in the first place a question of depth of penetration and efficiency: the higher the frequency, the less is the depth of penetration and the greater the efficiency of the work coil; above the frequency $f_{\min}$, whereby the depth of penetration is about one-eight of the diameter of the workpiece, there is, however, hardly any further increase in efficiency. Only when an even smaller depth of penetration is favourable for the result of the heat treatment does it serve any purpose to choose a frequency higher than $f_{\min}$. Now the smaller the diameter of the workpiece (imagined to be cylindrical) the higher is this minimum frequency $f_{\min}$. Consequently, both in the case of small workpieces and in cases where a small depth of penetration is desired, one has to use high frequencies which in practice can only be obtained by employing a valve generator; in other cases lower frequencies suffice, and rotary generators — though of a special construction — may be used.

The main consideration in choosing the frequency for capacitive heating is the rate of working, which has to be as high as possible. Now, as is to be seen from the formula (4), the rate at which the temperature rises is proportional to the frequency, so that in the first place one will aim at getting the highest possible frequency. Consequently valve generators are used exclusively. However, in order to be able to employ normal types of valves and to keep the efficiency of the generators high, not too high a frequency should be chosen, this being usually in the order of 10 Mc/s. The material constants, the permissible field strength and the temperature desired are generally such that with frequencies of this order the heating time is very short and a considerable saving is obtained compared with the time that would be required for any other method of heating.

In a valve generator for capacitive heating the capacitor, formed by the electrodes and the workpiece, may be a part of the resonant circuit. It has to be taken into account, however, that the loss resistance R (fig. 1*b*) may be very much smaller or greater than the optimum load resistance of the generator valve. *Fig.* 5 shows a system whereby the resistance R can be transformed to a favourable value by a suitable dimensioning of the capacitor C_1 and the inductor L_1.

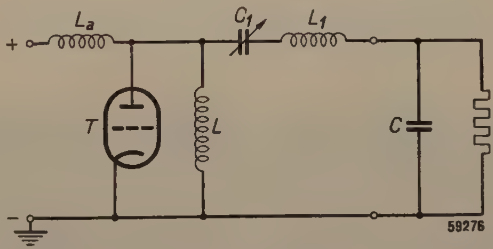


Fig. 5. Simplified circuit of a generator for capacitive heating. T = triode, C-R = equivalent circuit according to fig. 1*b*, L = inductance of the resonant circuit, C_1-L_1 = elements with which the resistance R is transformed to a value forming a favourable load for the triode, L_a = choke via which the anode is fed.

In this system there is apparently a resistance $R_{AC} = n^2R$ between the anode and the cathode of the generator valve, where the transformation ratio n is given by

$$n = \frac{L}{L + L_1} \cdot \frac{C + C_1}{C}.$$

With the aid of C_1 and/or L_1 it is therefore possible to give n any value greater or less than 1 which may be required to make R_{AC} equal the optimum load resistance of the valve.

Often only values of n greater than 1 are needed, in which case L_1 can be reduced to the self-inductance of the connecting wires and C_1 can be a variable capacitor.

Since the elements C_1 and L_1 bring about at the same time a transformation of the capacitance C, we have for the angular frequency generated :

$$\omega^2 = \frac{n}{LC}\,.$$

The maximum power generated in the workpiece with the Philips types of generators for capacitive heating lies between 0.3 and 50 kW. *Fig. 6* shows two generators, one for 2 kW and the other for 0.3 kW, in the special construction designed for the preheating of plastic pellets to be presently described. The generator illustrated in *fig. 7* has an output of 22 kW.

Applications of capacitive heating

The applications of capacitive heating, like those of induction heating, are still in course of development, but a number of fields have already been marked out where this method of heating offers striking advantages. Examples of such applications are the preheating of synthetic resins, the making of objects of bent plywood, the welding of certain thermoplastic materials, the twist-setting of rayon threads, etc.

We shall confine our considerations here to the first two applications (manufacture of plastics and the wood-working industry).

Capacitive heating in the manufacture of plastics

Phenoplasts (like "Philite") are thermo-setting substances, i.e. solids which when heated first become plastic but after a time harden and remain so upon cooling down, even if they should be heated again [4]). An article made of "Philite", for instance, may be manufactured in the following way. A weighed quantity of resins prepared from phenol and formaldehyde, mixed with fillers such as wood flour, mica and the like, is placed between the two halves of a steel mould. The mould is heated to about 160 °C, whilst at the same time the two halves are pressed together hydraulically. The moulding mass first becomes plastic, owing to the high temperature, and under the great pressure assumes the shape of the mould, after which it ultimately hardens.

The time that the mass has to be kept in the mould is mainly determined by the poor thermal conductivity of the mass. With a view to reducing

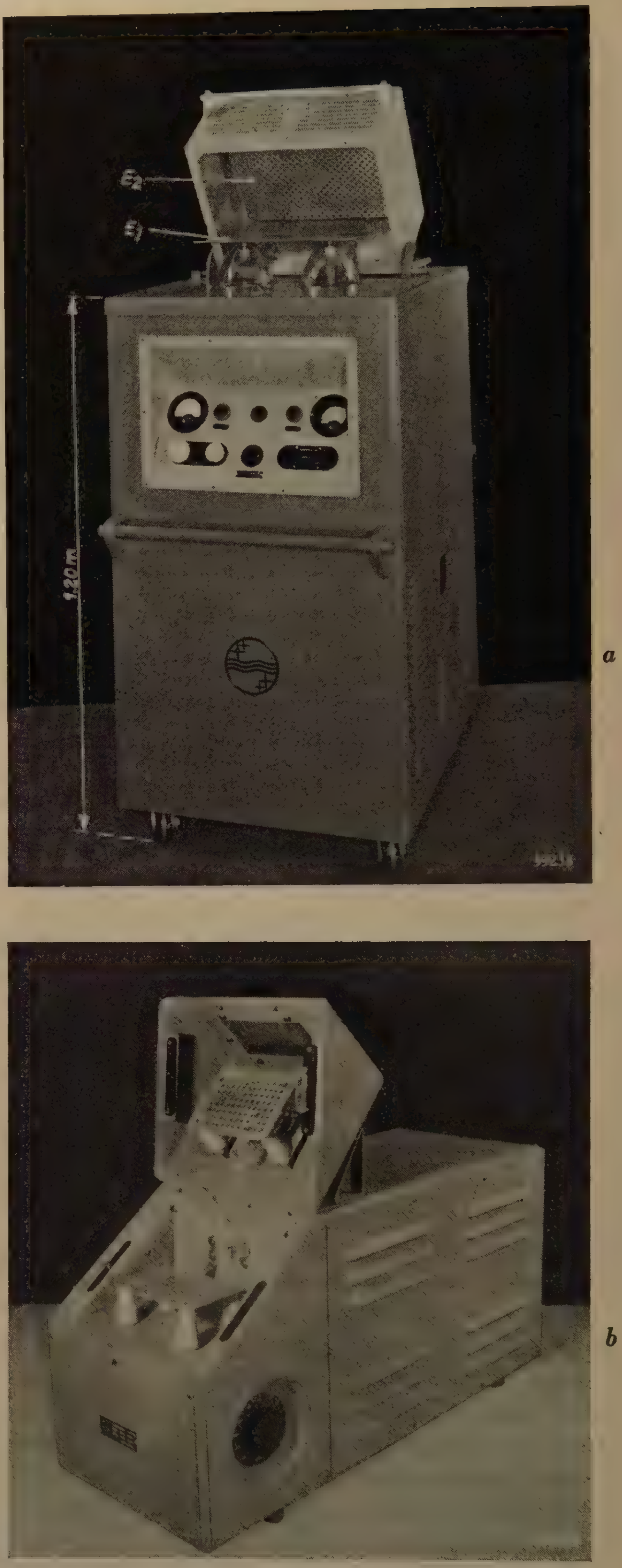


Fig. 6. *a*) Generator type SFG 136/21 for preheating plastic pellets, these placed on the plate E_1, which forms one of the electrodes. The other, earthed, electrode is the cap E_2, which is shown folded back. Frequency approx. 15 Mc/s. Maximum output 2 kW.

b) Generator for the same purpose for an output of 300 W, to be released shortly.

---

[4]) See for instance J. C. Derksen and M. Stel, Plastics and their application in the electrotechnical industry, Philips Techn. Rev. 11, 33-41, 1949 (No. 2).

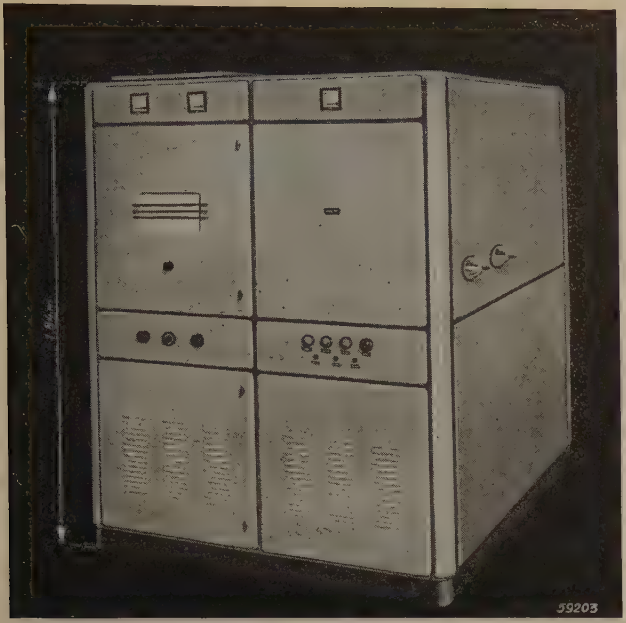


Fig. 7. Generator type SFG 134/01. Frequency approx. 2.5 to 8 Mc/s. Maximum output 22 kW.

the moulding time and thus increasing the output of a press per hour (or else reducing the number of presses required) it was very soon decided to preheat the mass before placing it in the mould. The idea was that if the mass could be preheated to a temperature just below that at which it begins to set, then only little additional heat would be needed in the mould. In this way much time would be gained in the moulding process, especially in the case of products with thick walls, which without preheating would take a long time to heat to the right temperature everywhere in the mould.

Now the conventional method of preheating, in an oven or furnace, is inefficient. When heat is applied externally the temperature on the outside of this poorly thermal-conducting moulding mass is much higher than that inside it, especially when it is desired to heat the mass quickly to a temperature approaching that of the mould. The danger then exists that after having become plastic the mass already begins to harden on the surface. To avoid this risk preheating is carried out very slowly and not beyond a rather low temperature, say 80 °C, but then, of course, it takes a long time and very little is gained in the pressing time. Moreover, in this way the mass does not reach that degree of plasticity that it does when it is preheated to a higher temperature and which has great advantages in the moulding process, since the more plastic the mass the less is the mould subject to wear and the lower is the pressure required (lighter presses suffice).

With capacitive heating the conditions are much more favourable, because the heat is generated inside the object itself. Owing to a certain amount of radiation the temperature on the outside of the object is always somewhat lower than that below the surface, but by means of thermal insulation or by keeping the electrodes themselves hot it is quite easy to ensure a highly uniform distribution of heat.

With the modern methods of manufacturing articles from "Philite" and similar plastics the mass is made in the form of pellets. A number of these pellets just sufficient to make the article that is being produced are placed in the mould after first being preheated. This preheating, to make the pellets plastic (see *fig. 8*), is done, for instance, with a generator like that illustrated in *fig. 9* or *10*. With the latter generator, for example, a charge of 100 grammes of artifical resin reaches a temperature of 110 °C in 1 minute. Once the generator has been running for a time the electrodes have become so heated by the charges already passed through that for the following charges very little heat is lost and thus a highly uniform distribution of heat is obtained. It is due to this that the pellets become plastic without showing any trace of the granulation that usually occurs when preheating in an oven. It is only with capacitive preheating that the aforementioned advantages of reduced

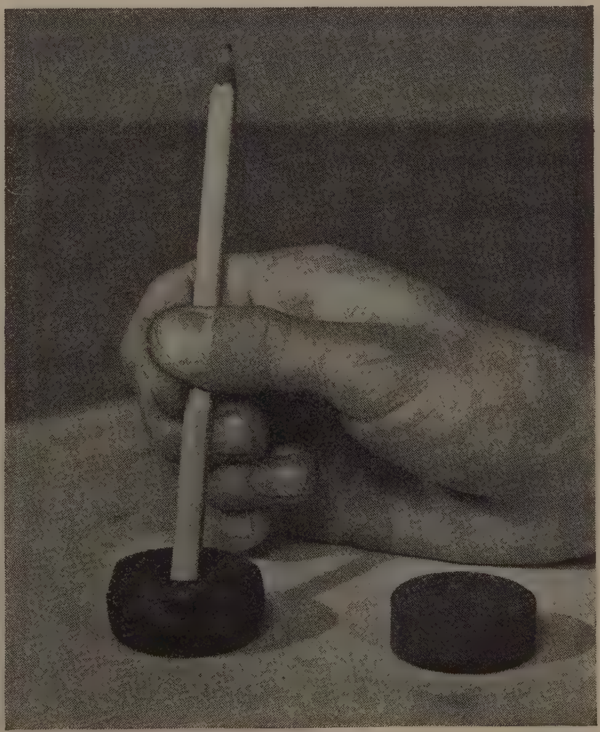


Fig. 8. Two thermosetting plastic pellets. The left-hand one has been made plastic by preheating and is ready for moulding.

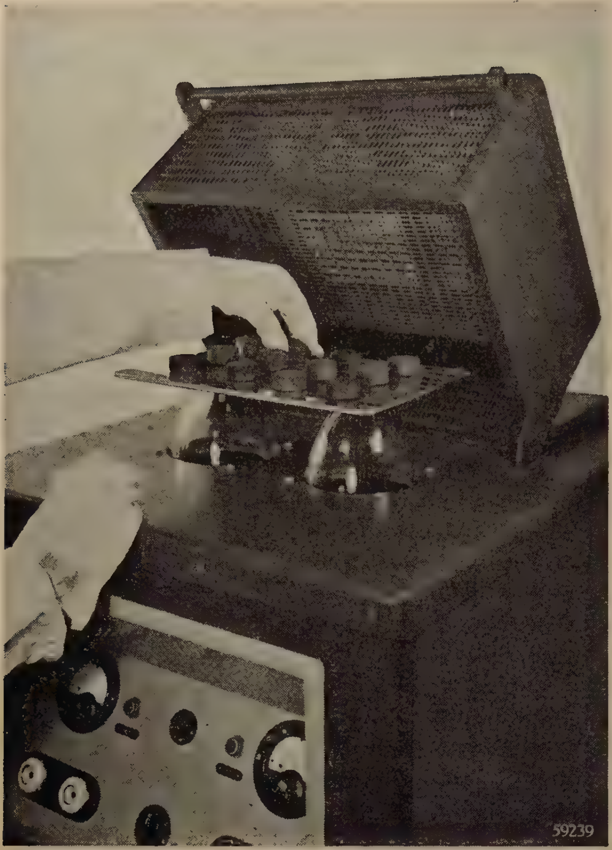


Fig. 9. Generator of the type illustrated in fig. 6*a*, with some preheated plastic pellets.

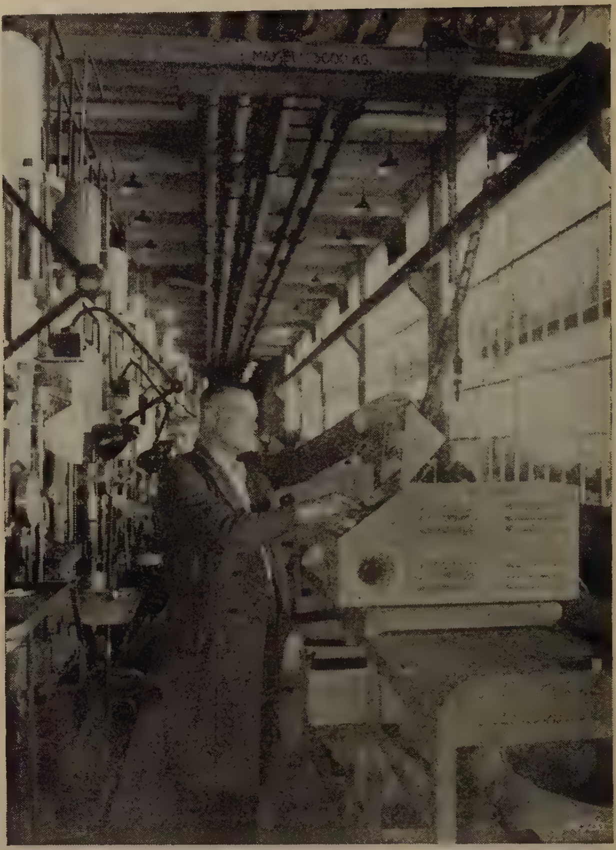


Fig. 10. Generators of the type depicted in fig. 6*b*, used in the "Philite" works at Eindhoven.

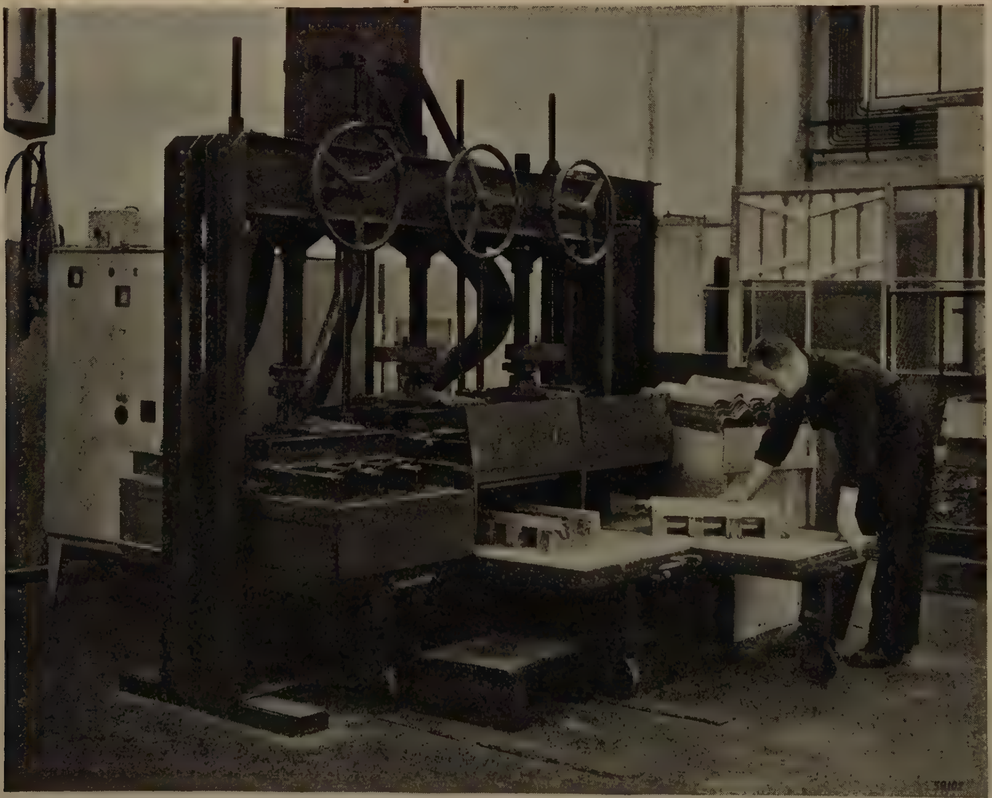


Fig. 11. Triple press for turning out corrugated sheets of plywood (seen on the right of the press). Two of the wooden moulds in which the sheets are pressed are shown. On the left in the background is the generator.

wear of the moulds and lower pressures are fully derived. Furthermore, if at the same time metal parts have to be moulded into the object the fact that the mass is made plastic throughout reduces the number of rejects; in a less plastic mass the metal parts are apt to get bent or be broken.

Compared with heating in an oven, capacitive heating has the further advantage that there is practically no inertia of the heat source, so that the rise in temperature can be stopped immediately, by switching off the generator (by hand or automatically). Thus it is easy to avoid heating to too high a temperature.

Consequently capacitive heating leads not only to the shorter moulding time mentioned but also, as follows from the foregoing, to a better control over all sorts of factors determining the quality of the product. Generally speaking it is therefore found that plastic objects produced by a manufacturing process where capacitive preheating is applied have better mechanical and chemical properties than those produced from a moulding mass preheated in some other way.

Capacitive heating in the wood-working industry

Plywood is made from a number of leaves of glued veneer stacked one upon the other. The glue used may be either urea or phenol formaldehyde with the addition of hardeners. Upon the stack of leaves being heated to a certain temperature (90 or 130 °C according to the kind of glue) and brought under pressure the hardeners bring about a polymerization of the glue which causes it to set hard. The plywood thus obtained is then ready for further processing (sawing, drilling, varnishing, etc).

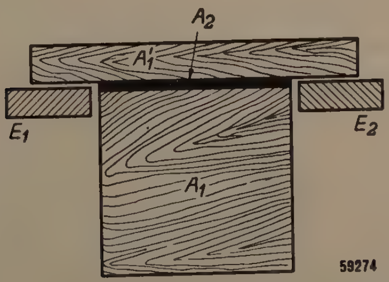


Fig. 12. A wooden block A_1 has to be glued onto a small plank A_1' with the glue A_2. The electrodes E_1, E_2 are applied in such a way as to obtain selective heating in favour of the glue.

By pressing in a shaped mould the plywood can be given all sorts of profiles, such as are often used for modern furniture, radio cabinets, etc.

The pressure needed in this process is so much lower than what has to be applied in the manufacture of plastics that no expensive steel moulds are required, simple wooden moulds answering quite well. This makes it possible for the capacitive heating to be carried out in a simple way during the moulding (in a period of time of the order of 60 seconds); all that is needed is two metal plates on the inside of the mould to act as electrodes.

Fig. 11 shows an installation for turning out corrugated sheets of wood as seen on the right of the press. This being a triple press, the generator can be run continuously.

In the manufacture of plywood the wood and the glue are connected in series. The material constants are such that no selective heating occurs. Selective heating can be obtained, however, in the gluing of larger workpieces, as we have seen in fig. 4. Another example of selective heating is the

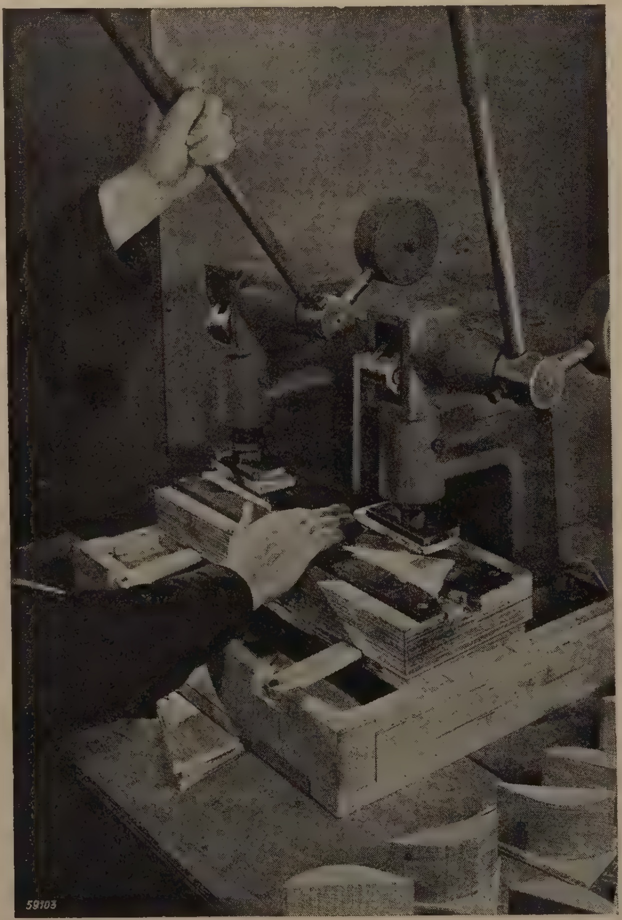


Fig. 13. Gluing by the method sketched in fig. 12. In the double hand press are to be seen the (triangular) planks; the blocks that are to be glued on are underneath the planks. The electrodes are out of sight.

gluing of a wooden block on a small plank (*fig. 12*). In order to direct as much of the energy as possible into the glue it has been purposely arranged to get, as it were, a parallel circuiting of the glue and the wood. Since the factor $(\varepsilon_r \tan \delta)/\varrho c$ of the glue is several times larger than that of the wood the heating is strongly selective in favour of the glue (cf. formula (5)). *Fig. 13* shows how this gluing is done with the aid of a simple hand press.

Summary. Capacitive heating is based upon the dielectric losses occurring in a dielectric situated in an alternating electric field. A formula is derived for the amount of heat generated per unit of time in a homogeneous dielectric. It appears that two simple cases of a non-homogeneous dielectric can be regarded as the parellel and series connection, respectively, of two homogeneous dielectrics; if the dielectrics are respectively glue and wood then with parallel connection an often desired selective heating may take place. After a brief consideration of the generators for capacitive heating, some important applications are discussed; manufacture of thermosetting plastics and of plywood, and the gluing of wood.

APPARATUS FOR PREPARATION OF METALS WITH AN EXACTLY KNOWN CONTENT OF IMPURITIES

by J. D. FAST.

621.365.56:66.046.516:669.77/.78

The presence of impurities in technical iron and steel determines for a large part the properties of these materials. It is difficult to study the influence of any particular element exactly because in nearly all cases several elements are present simultaneously. Only after thoroughly purifying the iron and then adding the desired element alone in an accurately weighed quantity is it possible to obtain data indicating the effect of that element.

Impurities in iron and steel

All normal technical steels and iron contain carbon, oxygen, nitrogen, sulphur and phosphorus. In special cases the last four constituents may, in combination with other elements, give the steel certain desired properties, but in most cases they are to be regarded as undesirable impurities which can never be entirely avoided in the large-scale production of steels. As a rule they have a particularly adverse influence upon the mechanical properties of the metal. The common grades of steel usually contain an amount of each of these four elements lying between 0.01 and 0.1 wt %. (In this paper all percentages are by weight.) In certain cases injurious effects may also be ascribed to carbon, which is not usually classed as an impurity and occurs in the various kinds of steel in greatly varying quantities (at least several hundredths per cent).

The harmful effect of these elements may become clear immediately after the shaping of the metal (by forging, rolling, drawing, etc) and the accompanying thermal treatment. Sometimes it also happens that the mechanical properties gradually change in the adverse sense subsequent to the last thermal or mechanical treatment, for instance while the metal is stored. One then speaks of ageing phenomena.

The question as to which impurity is actually responsible for a certain injurious effect that has come to light is often difficult to answer. The reason for this is that as a rule experiments are carried out with technical steels containing all the above impurities at the same time.

Probably the influence of sulphur is best known. In solid iron this element is practically insoluble but it combines with part of the iron, giving the chemical compound iron sulphide (FeS). In the liquid state iron and iron sulphide are homogeneously miscible in all proportions and form a eutectic with a comparatively low melting point (985 °C). When a sulphurous iron solidifies a metal is obtained in which the iron sulphide is located at the crystal boundaries. As a consequence in the mechanical treatment (e.g. forging) fractures occur along the crystal boundaries, both at low temperatures, because iron sulphide is brittle, and at high temperatures, owing to the presence of a thin layer of a liquid eutectic at the crystal boundaries. In order to counteract this injurious effect of sulphur an excess of manganese (with respect to the sulphur) is always added to the common technical steels, thereby binding the sulphur in the form of manganese sulphide. This compound has a higher melting point than iron and after solidification is found in the iron in the form of numerous small globules. Inclusions in such a form are comparatively harmless in metals.

Phosphorus cannot have the same effect as that described for sulphur because it is highly soluble in solid iron even at room temperature (about 1%).

In order to study the effect of carbon, oxygen or nitrogen separately an apparatus has been designed which makes it possible to add exactly known quantities of these elements to molten pure iron. This apparatus will be described here.

The properties of the alloys prepared in this way will be discussed in a subsequent article, where attention will particularly be paid to the notched-bar impact strength of the material and to the ageing phenomena already mentioned.

The melting apparatus

In order to obtain a metal of known and reproducible composition the melting is done by induction heating. Here only a brief description of this heating method will be given; it has been dealt with at length in an article by E. C. Witsenburg recently published in this Journal [1]).

The energy is applied to the material to be melted by generating in the melting chamber a high-frequency alternating magnetic field. The heat developed in the metal is derived mainly from the eddy currents set up in it. In the case of iron, with which we are specially concerned here, below the Curie point the hysteresis losses also play a part. The eddy currents occur (when very high frequencies are applied) only in the outer layers of the material. The field is generated by placing around the melting chamber a coiled copper tube through which the high-frequency current flows and which is kept cool with running water. The high-frequency generator used has an output of 15 kW at a frequency of 330,000 cycles per second. The coil has 20 turns and a diameter of 13 cm (5″).

---

[1]) Philips Techn. Rev. **11**, 165-175, 1949 (No. 6).

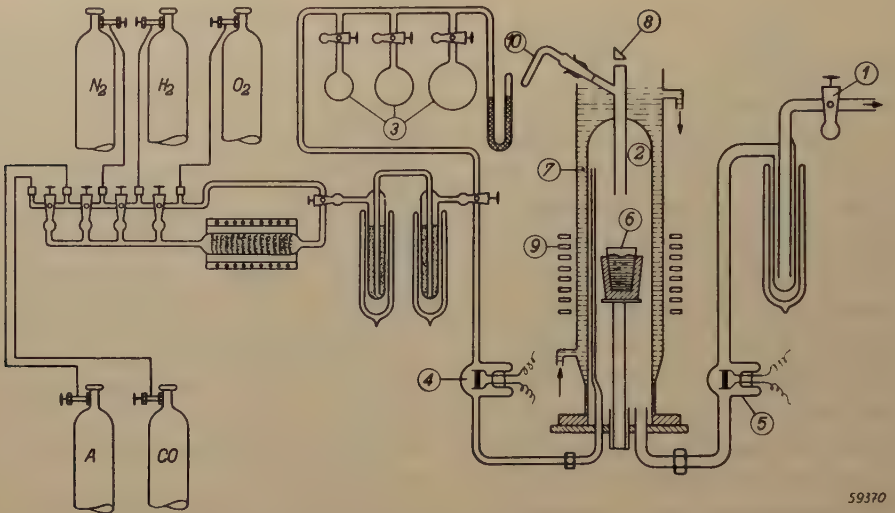


Fig. 1. Diagrammatic representation of the melting apparatus. *1* cock forming the connection with the high-vacuum pumps (not shown), *2* melting chamber, *3* bulbs with measured quantities of oxygen, *4* and *5* valves for detecting the presence of traces of water vapour and oxygen, *6* crucible, *7* cooling jacket, *8* prism for measuring the temperature of the melt with an optical pyrometer, *9* high-frequency coil, *10* rotatable tube for adding small quantities of solids to the melt. The diagram also shows a tube filled with copper turnings and placed in an oven, and two cooling vessels filled with the same material, serving to remove traces of oxygen and water vapour.

One of the great advantages of such an induction oven is that the melting can be carried out in an apparatus made entirely of glass, the walls of which can be cooled with running water. Thus both the influence of the atmosphere and that of gases released from the walls is almost entirely precluded 2). The apparatus can be evacuated up to a pressure of 10^{-8} atm; if desired it can then be filled with different pure gases. The crucibles are made of pure aluminium oxide. For some purposes crucibles of zirconium oxide or beryllium oxide are more suitable, but aluminium oxide is much cheaper.

Fig. 1 gives a diagrammatic representation (not drawn to scale) of the apparatus employed. On the extreme right is the cock (*1*) by means of which the melting chamber (*2*) can be connected to the high-vacuum pumps (not shown in the drawing). On the extreme left are cylinders containing nitrogen, hydrogen, oxygen, argon and carbon monoxide, which gases can be let into the melting chamber. Oxygen is admitted only in measured quantities, by using bulbs of known volumes (*3*). In the inlet pipe and also in the outlet pipe of the melting chamber are small valves, (*4*) and (*5*) respectively, with a ribbon of 18% Cr-8% Ni steel which can be heated by passing an electric current through it and which enables traces of oxygen or water vapour in hydrogen, nitrogen or argon to be detected by heat-tinting 3).

The crucible of Al_2O_3 (*6*) is contained in a larger crucible of clear quartz glass, the space between these being filled with coarse pieces of Al_2O_3 (fragments of discarded crucibles). The temperature of the molten metal can be measured with the aid of an optical pyrometer via a prism (*8*). A photograph of the whole set-up is given in *fig. 2*.

The preparation of metals of an exactly known composition

In order to produce iron with a definitely known content of impurities one has to start with the purest carbonyl iron. This is 99.9% iron and the

---

2) In this connection it may be interesting to point out that there is another method of achieving this object, namely by using concentrated solar energy by means of mirrors or lenses. See, e.g., F. Trombe, Utilization of solar energy, Research **1**, 393-400, 1947 (No. 9).

3) G. W. Rathenau and H. de Wit, Metallurgia **40**, 114, 1949.

impurities consist almost exclusively of carbon, oxygen and nitrogen. The nitrogen can be fully extracted by repeated melting in a good vacuum, whilst the same applies to the carbon if prior to the melting there was a large excess of oxygen in the metal; if that is not already the case a known quantity of oxygen can be admitted to the metal from the bulbs (*3*). When all carbon has been driven out in the form of CO then the excess of oxygen is removed by reducing the liquid metal with flowing hydrogen for a long time. From the tinting of the Cr-Ni steel ribbon (*5*) it can be seen whether the oxygen has been completely removed. The hydrogen is removed in turn by melting in a good vacuum or in flowing argon.

The pure iron obtained in this way contains no more than 0.001% of carbon + oxygen + nitrogen, and the desired elements may then be added to it. Solids can be added to the melt by turning the branch tube (*10*) (fig. 1). Carbon, if desired as an impurity, can be added to the liquid iron in this way in the form of pure graphite. If the substances to be added evaporate rather easily, as for instance manganese, then the addition is made in pure argon instead of in vacuum. The additions blend homogeneously with the iron in a short time as a result of the circulation in the bath set up by the electromagnetic forces [4]).

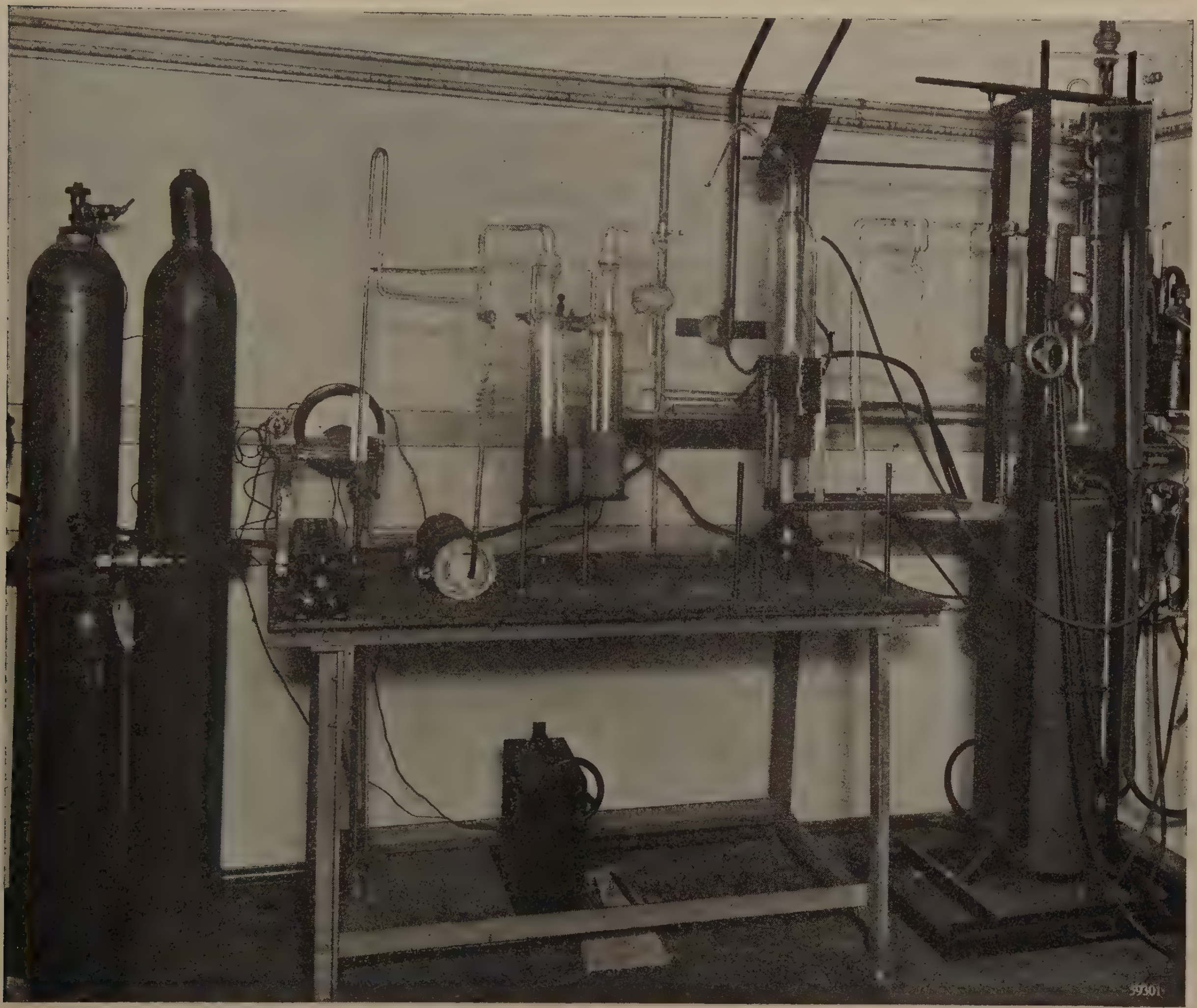


Fig. 2. Photograph of the set-up. On the extreme right are the high-vacuum pumps.

Oxygen and nitrogen are added in the gaseous state. The equilibrium pressure of oxygen over oxygen dissolved in liquid iron is negligibly low, so that the oxygen added is entirely absorbed by the iron. The equilibrium pressure of nitrogen over nitrogen in liquid iron, on the other hand, is high, so that only a relatively small part of the nitrogen added is absorbed. Moreover, as already discussed in this Journal [5]), upon the metal solidifying it gives off a large part of the absorbed nitrogen. As a consequence the metal swells and after solidification shows porosity and large cavities.

---

[4]) See, for instance, Philips Techn. Rev. 1, 59, 1936.

In order to obtain nitrogen-containing iron without porosity one proceeds as follows. While the metal is liquid (temperature about 1600 °C) the nitrogen pressure is reduced from 1 to 0.2 atm, thereby reducing the equilibrium concentration in the metal from 0.046 to 0.021% of nitrogen. The high-frequency current is then switched off and the nitrogen pressure immediately raised again to 1 atm (before solidification begins). The metal then solidifies to a compact mass and after cooling down contains 0.022 to 0.030% of nitrogen. This effect has been found experimentally and can be understood, to a certain extent, from the solubility curve for nitrogen in iron. Smaller contents are obtained by choosing lower nitrogen pressures.

The procedure described is highly suitable for the preparation of comparatively small quantities of metal (300 to 400 grammes). When preparing the largest quantities that can be produced in the apparatus (about 2 kg) two difficulties arise. In the first place it takes too long to remove the last 0.01% of oxygen from the molten metal exclusively with the aid of hydrogen. For large melts, therefore, this last amount of oxygen is chemically bound by adding to the liquid metal the required quantity (determined by calculation) of pure zirconium, pure titanium or pure aluminium through the tube (*10*). The ZrO_2, TiO_2 or Al_2O_3 thereby formed remains in the iron for the greater part in a finely divided state. The influence of the oxygen is then much less than when it is bound in the form of FeO. The second difficulty lies in the fact that pure iron absorbs nitrogen so very slowly that it is almost impracticable to charge 2 kg of iron with nitrogen in the relatively deep crucibles used. On the other hand nitrogen is more rapidly absorbed in molten oxygenated iron and still more rapidly in molten carbon-containing iron. For experiments where an amount of material larger than, say, 1 kg is needed (for instance for measuring the impact strength as a function of temperature) the influence of nitrogen upon pure iron cannot, therefore, be studied directly but has to be deduced from the behaviour of metals containing oxygen, carbon, oxygen plus nitrogen and carbon plus nitrogen.

Experiments have been carried out not only with carbonyl iron but also with mild steel as basic material, in order to conform more to practice. The procedure followed with mild steel is exactly the same as that applied with carbonyl iron.

In conclusion it may be observed that with the apparatus described here it is also quite possible to prepare alloys with higher percentages of foreign elements, such as magnet steels for example.

Summary. In order to study the influence of carbon, oxygen, nitrogen and other substances upon the properties of iron a melting apparatus has been designed by means of which the impurities can be added to extremely pure iron in exactly known quantities. Melting is done by means of induction heating; the melting process can be followed throughout on account of the crucible being contained in a cooled glass jacket. The preparation of the purest possible iron is described. In the case of relatively large quantities (more than about 1 kg) it is found impracticable to study the influence of nitrogen separately. This influence can then only be deduced from the differences in the behaviour of iron containing oxygen or carbon with and without nitrogen as admixture.

---

[5] J. D. Fast, Philips Techn. Rev. 11, 101-110, 1949 (No. 4).

INSTABILITY OF SPRINGS

by J. A. HARINGX.

621.272.2:620.166

Little attention has so far been paid to the phenomena of instability that are apt to occur with resilient elements, such as helical springs, rubber rods and flat spiral springs, when these are subject to great distortions. It appears possible to explain this so-called buckling theoretically. Some remarkable properties which are brought to light can be confirmed by simple experiments.

Introduction

The occurrence of phenomena of instability in the case of constructional elements, such as rods, plates and thin-walled cylinders, may be assumed to be generally known. With such elements it is characteristic that so long as the load does not exceed a certain limit they retain their original shape. When that limit is reached, however, the element becomes suddenly distorted to a practically unlimited extent, so that it no longer answers its purpose. Here this phenomenon will be referred to as buckling.

For the purposes of the following explanation it will be useful to investigate the case of a metal rod of circular cross section with the aid of some simple formulae. When a straight rod is loaded axially at first it remains straight, but upon a certain compressive force being reached it suddenly deflects laterally. The critical compressive force causing this buckling depends upon the rigidity of the rod with respect to bending and is given, according to Euler's formula, by:

$$P_{\mathrm{cr}} = \pi^2 \frac{EI}{l^2}\,, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where E is Young's modulus of elasticity of the material, l the length and I the moment of inertia of the cross section of the rod.

Although the changes in length of a metal rod are of course very small, for the sake of the analogy that is to be shown with the buckling phenomenon of helical springs a calculation will be made to determine in how far the rod is shortened before buckling takes place. The relation between the compression Δl and the compressive force P is given, as is known, by

$$\Delta l = \frac{Pl}{EF}\,, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where F represents the cross-sectional area of the rod. The relative compression immediately prior to the buckling thus amounts to

$$\xi = \frac{\Delta l}{l} = \frac{P_{\mathrm{cr}}}{EF}\,. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

On account of equation (1), and with $F = \pi D^2/4$ and $I = \pi D^4/64$ respectively, for a rod of diameter D this formula becomes:

$$\xi = \frac{\pi^2}{16}\left(\frac{D}{l}\right)^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The critical relative compression therefore appears to be dependent only upon the ratio l/D, which is to be construed as the slenderness of the rod, and not upon Young's modulus of the material.

The foregoing applies for a rod the ends of which are hinged or constrained parallel. For a rod with both ends clamped l has to be replaced in the formulae by half the rod length $l/2$, as may be seen from the following. *Fig. 1a* shows the deflection due to buckling in the case of a rod whose ends are constrained parallel. The rod is seen to assume the shape of a sine curve of half its wavelength. Fig. 1*b* shows a rod with the two ends clamped in the unstable state, the sine curve here consisting of a whole wave. Imagining this rod to be cut through the middle along the dotted line (*s*), it is seen that each half is in the same condition as the rod whose ends are constrained

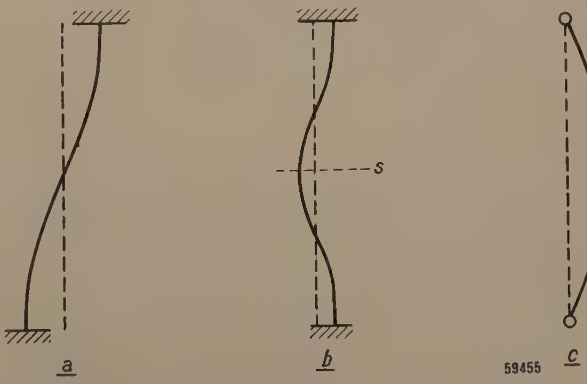


Fig. 1. The deflections that a rod undergoes when buckling under an axial force, *a*) for a rod the ends of which are constrained parallel, *b*) for a rod with clamped ends, *c*) for a rod with hinged ends. When the second rod is imagined as being cut in two along the dotted line *s* it is seen that the two halves are in the same state as the rod with its ends constrained parallel.

parallel. Thus here half the rod length determines the occurrence of buckling. The deflection due to buckling in the case of a rod with freely hinged ends is illustrated in fig. 1c, where the sine curve describes a half wavelength just as in fig. 1a.

It is to be noted that one must be careful when applying formula (4) because it holds only for elastic compressions of the rod (ξ smaller than about 0.001). In the event of the rod being subjected to such a load as to cause the yield point to be exceeded then considerable deviations arise.

It is obvious that under a compressive force helical springs may buckle just as well as straight rods. But in this respect, owing to their greater deformability, helical springs appear to possess some properties differing specifically from those of a straight metal rod. In what follows a description will be given of these remarkable properties as found by the author by calculation and subsequently confirmed experimentally. The buckling of elastic rubber rods is likewise briefly discussed. Further, attention will be paid to similar buckling phenomena shown by wound and unwound flat spiral springs, since it is also of importance to know under what conditions buckling takes place in these cases.

The behaviour of a helical spring under compression

Denoting the length of the spring in the unloaded state by l_0 and its diameter by D_0, in the case of a helical spring of circular wire section with its ends hinged or constrained parallel, the relative compression at which buckling takes place ($\xi = \Delta l/l_0$) amounts to:

$$\xi = 0.8125 \left\{ 1 \pm \sqrt{1 - 6.87\,(D_0/l_0)^2} \right\} . \quad . \qquad (5)$$

This formula [1]) has been derived by introducing a simplifying assumption: the helical spring is imagined to be replaced by an elastic prismatic rod which has to consist of such a material that not only its rigidity with respect to bending but also that with respect to compression and shearing correspond to those of the helical spring. The fact is that, owing to the great deformability of the helical spring, its rigidities with respect to compression and shearing also play a part, while these rigidities in the case of a straight metal rod can safely be regarded as being infinitely large. From this it follows that the calculation for the helical spring differs essentially from that for a rod.

The introduction of the rigidity with respect to shearing involves a complication in regard to the transverse force responsible for the shearing. As the author has demonstrated, one has to take the transverse force in that cross section of the elastic rod (replacing the spring) which in the unloaded state is perpendicular to the central line of the rod. With the method of calculation hitherto commonly used, however, the transverse force has been taken which acts in the cross section perpendicular to the central line in the loaded state.

The formula given also holds for the case where the ends are clamped provided l_0 is replaced by half the spring length $l_0/2$. If the cross section of the wire is rectangular then only the values of the coefficients change.

It is to be noted that for $l_0/D_0 > 5$ the smallest value of ξ in formula (5) can approximately be written as

$$\xi \approx 2.80\,(D_0/l_0)^2 , \quad . \quad . \quad . \quad . \quad . \qquad (6)$$

which shows a decided agreement with formula (4). Apart from the indices only the coefficients differ.

Just as in the case of the metal rod, Young's modulus of the material has disappeared from the formulae and the critical relative compression depends only upon the ratio of the length to the diameter and upon the manner in which the ends are fixed. With the helical spring, however, this is much more striking than in the case of the rod, because it means that neither the diameter of the wire nor the the pitch or number of coils has any effect upon the buckling phenomenon.

The relation, given in equation (5), between the critical relative compression ξ and the slenderness of the spring l_0/D_0 is graphically represented in *fig. 2*. The plotted curve is to be regarded as the line marking off the region of all combinations of spring slenderness and relative compression leading to instability. It shows that there is a certain critical slenderness ($l_0/D_0 = 2.62$ in the case where the

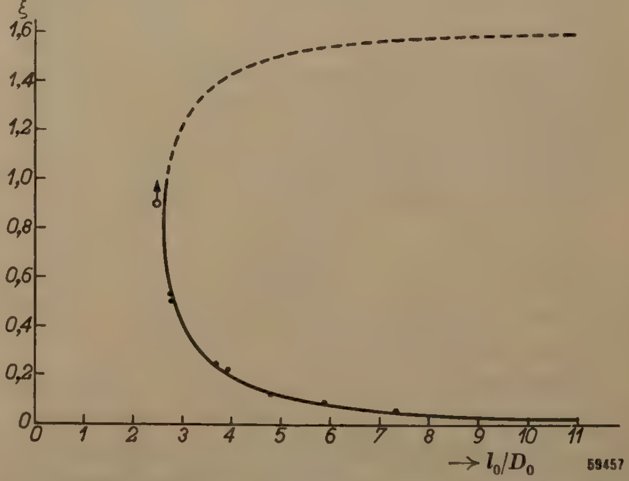


Fig. 2. The critical relative compression of a helical spring of circular wire section having ends hinged or constrained parallel, plotted as a function of the slenderness of the spring l_0/D_0 (see eq.(5)). The dots represent the results of measurements. For degrees of slenderness less than the value 2.62 no buckling takes place, even under full compression. The vertical arrow indicates that the respective spring did not buckle.

---

[1]) J. A. Haringx, On highly compressible helical springs and rubber rods, and their application for vibration-free mountings, Part I, sections 2 and 3, Philips Research Reports 3, 406-414, 1948 (No. 6).

ends of the spring are hinged or constrained parallel, and 5.24 if the ends are clamped) below which the spring does not buckle and above which it does. As a result of the steepness of the curve near this critical value, it so happens that any small variation of slenderness in the neighbourhood of 2.62 or 5.24 gives rise to very great differences in the behaviour of the spring. If this value of the slenderness is only slightly exceeded then there is a decided buckling of the spring when it is compressed to about half its original length. On the other hand there is no question at all of instability if the slenderness of the spring is slightly less than the critical value. This difference is demonstrated in *figures 3a* and *3b* by the two helical springs clamped at the ends. The right-hand spring with slenderness $l_0/D_0 = 5.3$ definitely buckles, whereas the left-hand one with slenderness $l_0/D_0 = 5.0$ remains straight and does not buckle.

The results of the calculation have been further checked experimentally by measuring the relative compression at which buckling takes place in the case of a number of helical springs of different degrees of slenderness ($D = 18$ mm, wire diameter $= 0.5$ mm and 2 mm, pitch $= 5$ mm). The results of these measurements are indicated in fig. 2 by dots. The arrow indicates that the spring of the corresponding slenderness did not buckle. As is to be seen, there is very satisfactory agreement with the calculated curve.

Under certain circumstances, however, there may be not inconsiderable deviations. In the first place the "clamping" of the ends of the spring is usually no fixture at all, but only a pressing of the flat-wound end coils against a flat surface. The elasticity of the end coils then plays a part, with the result that the spring buckles when the compression is apparently too small.

In the second place the length of the spring is understood to be the "free" length, that is to say, the length of that section of the spring taking part in the deflection. Now it may happen, especially if the pitch of the spring is comparatively small with respect to the thickness of the wire, that in the compressed state the last free coils make contact with the end coils, in which case buckling takes place at an apparently too large compression. It is possible to avoid these and other effects to a certain extent by giving the planes of the end coils in the unloaded state an oblique position. For further details reference is made to the calculation of the effect of the flat-wound end coils upon the lateral rigidity of helical compression springs [2]).

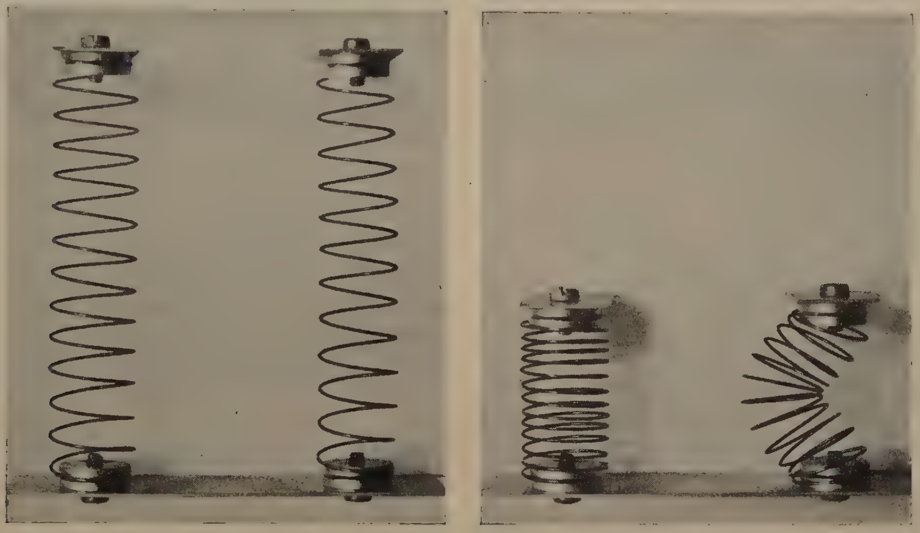

a *b*

Fig. 3. *a*) Two non-loaded helical springs with clamped ends and differing slightly in size. The slenderness of the left-hand spring is 5.0, thus a little below the critical value, which for springs clamped at both ends amounts to 5.24; the slenderness of the right-hand spring is 5.3, thus a little above the critical value. *b*) The same springs under a relative compression of 60 %. The left-hand one remains straight, the right-hand one buckles.

Behaviour of the helical spring at a compression greater than the original length

Normally a (cylindrical) helical spring can never be compressed beyond the point where all contiguous coils make contact. For this reason only the full-drawn part of the curve in fig. 2, corresponding to compressions Δl less than the original length of the spring l_0 ($\xi < 1$), is of practical importance.

---

[2]) See footnote[1]), Part II, sections 11 and 12, Philips Research Reports 4, 57-68, 1949 (No. 1).

Theoretically, however, the region $\xi > 1$ is very interesting, because, according to the dotted part of the curve , a spring that has once become unstable would become stable again if it were compressed far enough. To show that this is indeed the case we need a spring compressed farther than its original length, a condition which at first sight appears to be difficult of realization but which can in fact be brought about by turning a normal helical spring inside out.

To do this we take a helical spring with a small wire diameter, for instance 1/50th of the coil diameter. Beginning at one end of the spring, we draw the second coil over the first one, then the third coil over the first and second ones, and so on, thus reversing the positions of each coil's neighbours. Without our help they cannot return to their original position, so that they are of necessity pressed against each other. Consequently the spring is already under an initial load and subject to a compression Δl exceeding the original length of the spring l_0 by the product of the number of coils and the diameter of the wire. The relative compression is then indeed greater than 1. When such a spring that has been turned inside out is "stretched" (corresponding to a further compression of the original spring) then ξ is still further increased. Now this stretching is accompanied by a very remarkable phenomenon. The coils first slide along each other and tend to take up a parallel position; the spring assumes the shape of a flat ribbon, as shown by the middle part of the spring in *fig. 4a*. When, however, ξ exceeds a certain value the wire of the spring suddenly takes the shape of a helix (fig. 4*b*). Upon the spring being made "shorter" again, with more or less the same critical value of ξ the coils tumble over and the wire of the spring resumes the shape of fig. *4a*, thus proving that the spring does indeed show a second transition from the stable to the unstable state in the region where $\xi > 1$.

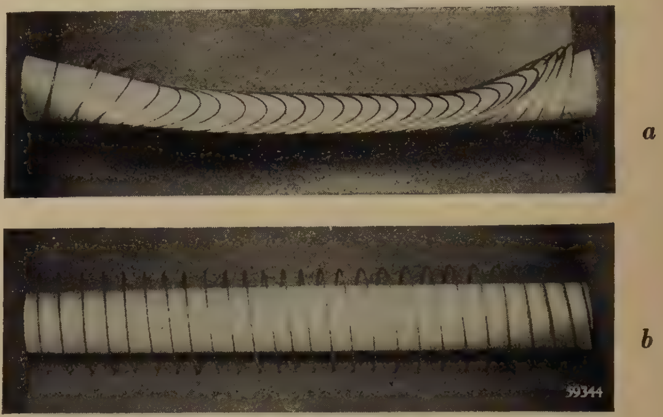


Fig. 4. *a*) A helical spring with clamped ends at a compression greater than the original length of the spring but not yet so great as to cause the second stability transition point to be passed. In order to give a better insight into the three-dimensional shape of the spring wire the spring has been passed over a thin paper cylinder, which does not constitute any hindrance for the free distortion of the spring. *b*) A helical spring with clamped ends at a compression greater than the original length of the spring, being so great that the second stability transition point is passed and the spring has again assumed the shape of a helix.

A further calculation [3]) has shown that the part of the curve in fig. 2 for $\xi > 1$ can only hold for

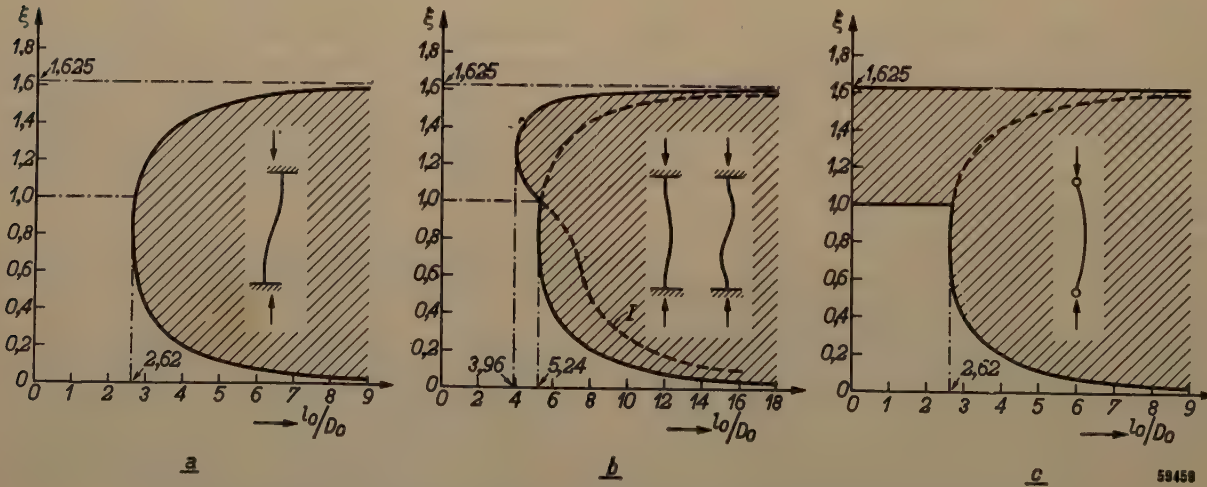


Fig. 5. The relation between the relative compression at which buckling takes place and the slenderness of the spring, for different methods of clamping the ends. *a*) Spring with its ends constrained parallel; this is the same curve as given in fig. 2; *b*) spring with clamped ends; here asymmetrical deflection takes place at values of $\xi > 1$; the curve relating to this deflection is that denoted by *I*; *c*) spring with hinged ends; the whole of the region between $\xi = 1$ and $\xi = 1.625$ corresponds to unstable states.

---

[3]) J. A. Haringx, Elastic stability of helical springs at a compression larger than original length, Applied Scientific Research, A1, 417-434, 1949.

helical springs with the ends constrained parallel. The region of the combinations of spring slenderness and relative compression leading in this case to instability is represented by the hatched part of *fig. 5a*. The drawing in the open space of the diagram represents a possible deflection of the central line of the helical spring and gives an idea of the manner in which the ends of the spring are fixed. In the case of springs with clamped ends — taking into account a factor 2 for the horizontal scale — the upper part of the curve in fig. 2 assumes a different shape; see fig. 5*b*. The abrupt change at $\xi = 1$ from one curve to the other is related to the variable preference for the symmetrical or the asymmetrical deflection of the spring. Springs with hinged ends are always unstable for all values of ξ between 1 and 1.625, as indicated in fig. 5*c*.

Qualitatively the experiments confirm the results of the new calculation for the various methods of fixing the spring ends, but quantitatively there are certain deviations. It must be borne in mind, however, that in the region of $\xi > 1$ the coils of the spring tumble over through angles of almost 90° (see fig. 4*a*), so that the rotations can no longer be regarded as being infinitely small, as was assumed in the calculation.

Behaviour of a rubber rod under compression

Although, for the sake of simplification, the buckling of helical springs has been calculated with the spring imagined as being replaced by a (fictitious) elastic prismatic rod, the calculation for an elastic rod actually realizable is much more complicated. This is due, inter alia, to the fact that the relation between the compression of the rod and the axial load is no longer linear. A calculation for a rubber rod of circular cross section (original length l_0 and original diameter D_0) having the ends hinged or constained parallel yields the result [4]):

$$\xi = \frac{1}{1 + 1.62\,(l_0/D_0)^2} \; . \; . \; . \; . \; . \quad (7)$$

Here again the critical relative compression $\xi = \Delta l/l_0$ depends only upon the slenderness l_0/D_0 of the rod. Where l_0/D_0 is greater than 5 formula (7) is practically identical with formula (4), as was to be expected.

The curve in *fig. 6* is a graphical representation of equation (7), whilst the dots plotted have been taken from measurements carried out by Kosten [5]).

The agreement between the calculation and the experiment can be said to be very satisfactory. Whereas, however, according to the calculation every rod, however short, must buckle, experimentally a certain critical slenderness is found below which there is no possibility of buckling ($l_0/D_0 \approx 0.6$; see the measuring point in fig. 6 with vertical arrow indicating that the respective rod did not buckle). This behaviour bears a strong resemblance to that of helical springs, except that in the latter case the calculation provided a direct explanation for this behaviour. It must not be forgotten, however, that the elastic properties of highly compressed rubber can only approximately be expressed in formulae. For compressions greater than 50 % these formulae are certainly no longer reliable, and it is for this reason that the calculated curve for values of ξ greater than 0.5 has been drawn with a dash line. Moreover, under compression rubber rods usually lose the prismatic shape owing to their ends being fixed, and a thin layer of the material at the ends has to be regarded as being absolutely incompressible. Although this effect too can be partly taken into account, it still remains a source of deviations.

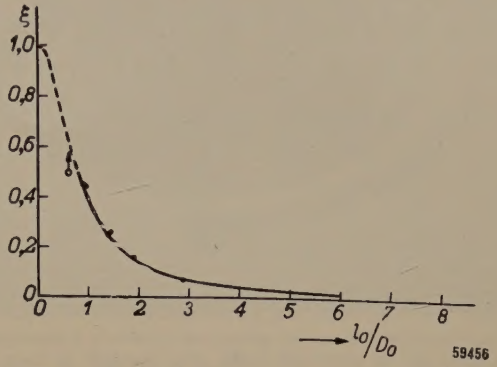


Fig. 6. The critical relative compression of a rubber rod of circular cross section and ends hinged or constrained parallel plotted as a function of l_0/D_0. The dots indicate the results of measurements by Kosten [5]). The arrow denotes that the respective rod did not buckle.

Phenomena of instability of a flat spiral spring

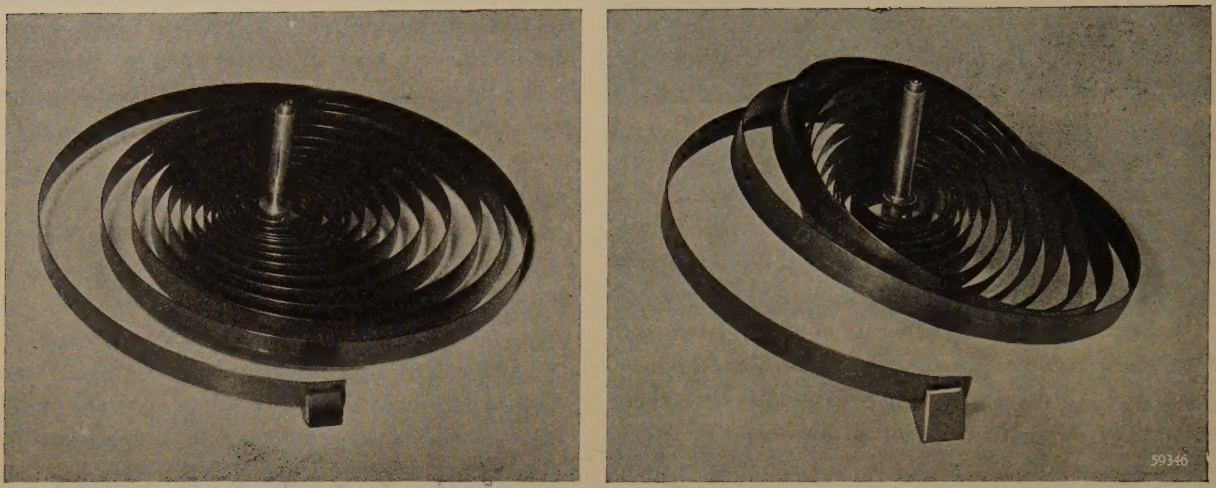
Apart from an axial compressive force, also a torque or a combination of these two loadings may cause a helical spring to buckle [6]). Since such cases, however, seldom occur in practice we shall not pay any further attention to them in this paper. Nevertheless this fact is mentioned because a relationship exists between the helical spring subjected to torsion and a flat spiral spring which has been wound up. This is readily understood when the length of the helical spring is imagined as being reduced to nil. It will therefore not be surprising that flat spiral springs, like those used in timepieces, are apt to buckle when being wound up or unwound, just as is the case with helical springs subjected to torsion.

---

[4]) See footnote [1]), Part III, sections 1 and 2, Philips Research Reports 4, 206-216, 1949 (No. 3).

[5]) C. W. Kosten, On the elastic properties of vulcanized rubber (in Dutch), Thesis Delft 1942, p. 66.

[6]) See footnote[1]), Part I, sections 8 and 9, Philips Research Reports 3, 435-449, 1948 (No. 6).



a *b*

Fig. 7. *a*) A flat spiral spring wound up just far enough to avoid buckling. *b*) The same spring after instability has set in.

When one of the ends of such a spiral spring is clamped and the other end is turned about the axis of the spiral then at first the coils continue to lie in the plane of the spiral as shown in *fig. 7a*, but as soon as the spring has been wound up a critical number of turns, the coils suddenly tilt in the manner indicated in fig. 7*b*. This critical number of turns has been calculated for spiral springs wound up and unwound [7]). In the case of spiral springs having a large number of coils the calculation leads to the following result. The number of coils that the spring has to be wound up or unwound before buckling takes place does not appear to depend upon Young's modulus of the material nor upon the number of coils and the dimensions of the spiral. The only determining factor is the ratio of the sides *a* and *b* of the (rectangular) wire section. The relation between this ratio and the critical number of coils *N* is represented in *fig. 8*. The side *a* of the wire section is directed radially and the side *b* is perpendicular to the plane of the spiral.

When b/a is greater than 10, as will nearly always be the case in practice, the influence of this ratio b/a disappears almost entirely (N varies only from 2.84 to 3.08). It may therefore be said that for most spiral springs the number of turns required to cause buckling is practically the same. This number of turns is approximately three.

As a rule the number of coils of a spiral spring will be rather small and the question is whether the results found also hold for such a spring. For spiral springs with a small number of coils the calculation is much more complicated, but it can nevertheless be carried out for a (fictitious) spiral spring having circular coils of equal diameter. It appears that a distinction has to be made between the number of turns the spring has to be wound up and the number of turns it has to be unwound to cause buckling. The differences with respect to the results for spiral springs having a large number of coils are, however, small.

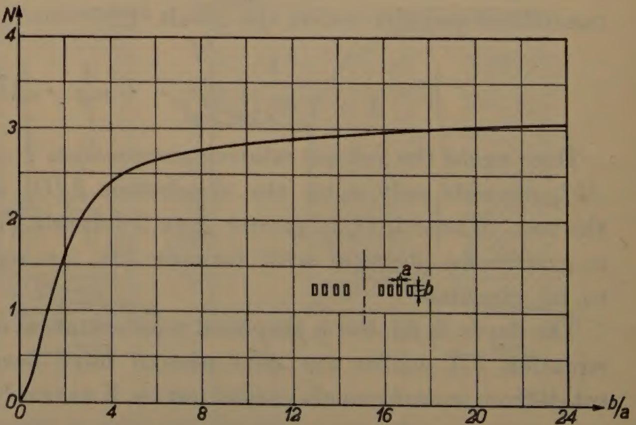


Fig. 8. The relation between the number of turns N that a flat spiral spring with (infinitely) large number of coils has to be wound up or unwound to cause it to buckle, and the ratio of the sides b/a of the rectangular wire section.

---

[7]) J. A. Haringx, Elastic stability of flat spiral springs, Applied Scientific Research, A2, 9-30, 1949.

To verify the results of the calculations tests were carried out with some spiral springs with different numbers of coils and different wire sections to determine at what number of turns they buckle. The results are given in the table below, where a comparison between the third and fourth columns shows that there is satisfactory agreement with the calculation.

The number of turns made in winding up the spring has been taken as positive, so that the negative sign on the last line in the table means that the spring was unwound. For judging the effect of the number of coils it is indicated in the last column what number of turns would be critical for spiral springs with (infinitely) large number of coils. The difference does indeed appear to be small.

It is to be noted that during the experiments the clamped ends of the springs were so arranged that successive coils could not touch each other and the spring remained flat until the moment that it suddenly buckled. If these steps for a careful adjustment of the position of the wire ends are not taken the spring will jump out of its plane at a 10 to 20 % smaller number of turns.

Number of coils	Ratio of sides of the wire section b/a	Number of turns required to cause buckling		
		Measured	Calculated	Calculated for infinite number of coils (fig. 8)
3.1	10	2.5	2.52	2.84
4.2	8.6	2.6	2.55	2.78
4.5	10	2.6	2.62	2.84
13.5	19	2.75	2.89	3.00
13.5	19	−2.9	−3.10	−3.00

Buckling can be avoided by placing the spiral spring in a box or between two parallel discs, but then there is inevitable friction between the spring and the wall or between the coils themselves, with the result that the magnitude of the torque of the spring is not reproducible. This objection can usually be accepted, but it is quite possible that in certain cases a frictionless construction is to be preferred. For instance, friction must certainly be avoided if the spiral spring is to be used as a resilient element in a controlling or measuring apparatus which has to be rotatable over large angles. When only one spiral spring is used the angle of rotation is then limited to the maximum of three turns mentioned above.

Summary. Some properties of helical springs, rubber rods and flat spiral springs (as used in timepieces) are discussed in their relation to buckling. The analogous phenomenon in the case of a metal rod is dealt with as an introduction to the discussion of the behaviour of a helical spring. The relative compression at which the spring of circular wire section buckles is found to depend only upon the ratio of the length to the diameter of the spring (its so-called slenderness) and upon the manner in which its ends are fixed. There is a minimum slenderness below which the spring does not buckle even in the fully compressed state. A second transition from the stable to the unstable state occurs at a compression greater than the original length of the spring. Such a compression can be brought about experimentally. Also in the case of the rubber rod the buckling is governed by the ratio of the length to the diameter and the manner in which the ends are fixed. In the case of a flat spiral spring the angle over which this has to be wound up or unwound to cause buckling depends only to a very small degree upon the dimensions of the spring; it appears to be impossible to design a spiral spring which can be wound up more than three turns without buckling.

ABSTRACTS OF RECENT SCIENTIFIC PUBLICATIONS OF THE N.V. PHILIPS' GLOEILAMPENFABRIEKEN

Reprints of these papers not marked with an asterisk can be obtained free of charge upon application to the Administration of the Research Laboratory, Kastanjelaan, Eindhoven, Netherlands.

R 112: G. W. Rathenau and J. F. H. Custers: Secondary recrystallization of face-centred Ni-Fe alloys (Philips Res. Rep. 4, 241-260, 1949, No. 4).

The primary texture of severely rolled Ni-Fe alloys is the cubic orientation. At high temperatures abnormal grain growth may occur: secondary recrystallization. The orientations of secondaries are different from the cubic orientation. Several new orientations were found. Alike orientations are encountered, whether the secondaries grow in a normal way or under quite different conditions (1-3 below). Normal secondary recrystallization is supposed to be the grain growth of primary crystals with a high temperature of primary recrystallization. One is led to this view by the following experimental evidence: (1) Thoroughly annealed strips were locally deformed by a pin prick and reheated. Large crystals grew frequently from the pinhole into the cubic matrix. Pin pricks made in secondary crystals did not give rise to grain growth outside the area of local deformation. (2) By intermediate local annealing a soft region was inserted in a cold-rolled strip. On recrystallization large secondaries grew from the boundary of the soft region into the cubic matrix. (3) Besides rolling, a deformation of another kind, such as cutting, hammering, was locally applied. Recrystallization produces secondaries which grow from the boundary of the locally deformed area into the cubic matrix. On cutting rolled strips, the direction of cutting and the orientation of the resulting secondaries proved to be related. (4) Rapid heating in the temperature region of primary recrystallization favours the nucleation of secondaries. (5) Large secondaries were observed only within a restricted region of rolling deformation, the extension of which depends on the rate of heating.

R 113: J. A. Haringx: On highly compressible helical springs and rubber rods, and their application for vibration-free mountings, IV (Philips Res. Rep. 4, 261-290, 1949 No. 4).

In this paper (see Abstracts Nos **R 94, R 101** and **R 109**) the different types of vibration-free mounting are treated, though in the one-dimensional case only. Of the well-known elementary systems with relative or absolute damping the latter is superior, since the relative damping unfavourably affects the forced vibrations at frequencies exceeding $\sqrt{2}$ times the resonant frequency. Absolute damping can unfortunately not be realized, but it can be imitated by applying an auxiliary mass in addition to the damping element. Usually the application of such a damped dynamic vibration absorber is limited by the demand that the auxiliary mass shall be small in relation to the main mass; with vibration-free mounting of instruments, however, there is in general no objection to the auxiliary mass being so large as to be able to take full advantage of the damping properties of this system in the region of the resonances. For optimum results some constructional conditions in respect of the spring rigidities and the coefficient of damping have to be satisfied. To determine these, the forced vibrations are examined by means of frequency characteristics, whilst for the judgement of the decay of the free vibrations a fictitious logarithmic decrement is introduced. Cf. Philips Techn. Rev. **9**, 16-23, 85-90, 1947.

R 114: E. Labin: Théorie de la synchronisation par contrôle de phase (Philips Res. Rep. 4, 291-315, 1949, No. 4). (Theory of synchronization by phase control; in French.)

An auto-oscillator, the frequency of which is controllable by means of D.C. voltage, can be synchronized with a pilot by taking the control voltage from a mixing stage in which the phases of the the two oscillators are compared. This method is first discussed when both oscillations are sinusoidal, and then extended to the case where one of the two oscillations consists of pulses, a condition that has been realized in the so-called I.G.O. system (impulse-governed oscillator). The salient properties of this method of synchronization as revealed by experiment are explained.

R 115: A. Guinier and J. Tennevin: Comparison of the perfection of the crystals of primary and secondary recrystallization (Philips Res. Rep. 4, 316-318, 1949, No. 4).

In connection with experiments on secondary recrystallization of Ni-Fe alloys (see Abstract No. **R 112**) the degree of perfection of primary and secondary crystals has been investigated, using a method previously described by the authors. Secondary and primary crystals proved to be equally perfect, whereas in the case of the growth of large crystals of Al, preceding primary crystallization, as described by Chevigny, considerable imperfections were found. Considering their limited sensitivity the experiments do not contradict the opinions expressed by Rathenau and Custers.